

Journal of Buddhist Ethics

ISSN 1076-9005

<http://blogs.dickinson.edu/buddhistethics/>

Volume 19, 2012

# Buddhist Reductionism and Free Will: Paleo-compatibilism

Riccardo Repetti  
Kingsborough College, CUNY

*Copyright Notice:* Digital copies of this work may be made and distributed provided no change is made and no alteration is made to the content. Reproduction in any other format, with the exception of a single copy for private study, requires the written permission of the author. All enquiries to: [cozort@dickinson.edu](mailto:cozort@dickinson.edu).



# Buddhist Reductionism and Free Will: Paleo-compatibilism

Riccardo Repetti<sup>1</sup>

## Abstract

This is the second article in a four-article series that examines Buddhist responses to the Western philosophical problem of whether free will is compatible with “determinism,” the doctrine of universal causation. The first article focused on the first publications on this issue in the 1970s, the “early period”; the present article and the next examine key responses published in the last part of the Twentieth century and first part of the Twenty-first, the “middle period”; and the fourth article will examine responses published in the last few years. Whereas early-period scholars endorsed compatibilism, in the middle period the pendulum moved the other way: Mark Siderits argued for a Buddhist version of partial incompatibilism, semi-compatibilism, or “paleo-compatibilism,” and Charles Goodman argued for a straightforward Buddhist hard determinism. The present article focuses on Sider-

---

<sup>1</sup> Department of History, Philosophy and Political Science, Kingsborough College, CUNY.  
Email: [rrepetti@kingsborough.edu](mailto:rrepetti@kingsborough.edu).

its's paleo-compatibilism; the subsequent article focuses on Goodman's hard determinism.

### Conclusions of the Early-Period Scholarship<sup>2</sup>

In the first article in this series ("Earlier") I examined the writings of early-period Buddhist scholars Story, Rāhula, Gómez, and Kalupahana regarding the Western philosophical problem of whether free will is compatible with "determinism," the doctrine of universal lawful causation. These "early-period" scholars resist a straightforward equation of the Buddhist doctrine of dependent origination (*pratītya samutpāda*), which asserts the dependence of all conditioned/composite phenomena on previous or simultaneous impartite micro-phenomena, with either a "rigid" determinism or a "chaotic" indeterminism, opting for a "middle way" between both. Some hold that Hume's model of causation, as mere *constant conjunction* obtaining among pairs of contingent event types, provides a "middle way" for Buddhist compatibilism.<sup>3</sup> But Humean causation involves generalizations about *contingent* event pairings; determinism involves *necessary* ones. Thus, determinism cannot be Humean; and if dependent origination is Humean, it is not deterministic. The Buddhist reply to the free will problem would then be simple: determinism is false. However, that undermines the idea that dependent origination is causation, rather than an error theory about it (like Hume's projectivist account). The problem is that causation arguably does too much substantive work in Buddhism to be amenable to a simple error theory.

Early-period scholars seem not to have noticed another middle-path option: *soft* determinism. "Hard determinists" hold that determin-

---

<sup>2</sup> I would like to thank Dan Cozort, Mark Siderits, Claire Gaynor, and an anonymous reviewer at the *Journal of Buddhist Ethics* for comments to the present article.

<sup>3</sup> Siderits also suggests this Humean model in *Persons*.

ism is true, entails an invariable series of events, and thus whatever happens cannot have been otherwise, so we lack free will. “Soft determinists” reject only the conclusion (“so we lack free will”), for various reasons, because, for example, knowledge of cause and effect renders undesirable events evitable.<sup>4</sup> The Buddha emphasizes that knowledge of cause and effect (between beliefs, volitions, actions, and consequences) and the cultivation of mindfulness of beliefs, volitions, and actions are the basic means to liberation and the end of suffering. Thus, if dependent origination is deterministic, the form endorsed by the Buddha would arguably be *soft* determinism.

Early-period scholarship addresses, but leaves unresolved, the Buddhist nonacceptance of the self, which raises the question of how a nonagent could be autonomous. Middle-period scholars, however, address this question directly. Goodman rests his argument against free will on the claim that because Buddhism rejects the self it rejects any responsible agent. A related issue, unexplored in early-period scholarship but taken up by Siderits, is mereological (part/whole) reductionism, the idea that all conditioned phenomena that appear to be *substantive* wholes do not ultimately exist as such, that is, as anything beyond their aggregated parts. The issues are related because if the self is no more than an aggregation of psychophysical events/processes, it is unreal; hence, if free will rests on an unreal entity, free will is unreal. Whereas

---

<sup>4</sup> Fatalists believe that if a certain event is fated, no antecedent interference can prevent it. *Local* fatalists think *certain* events are fated; *global* fatalists think *all* events are. Whereas fatalism is *acausal*, determinism is *causal*: every event is necessitated by causal laws. Global fatalists and determinists agree on a single, invariable, necessary series of events, but for acausal and causal reasons, respectively. Most hard and soft determinists agree that the invariable series runs through causally effective choices, but disagree about whether agent-proximal activities (such as belief, volition, choice, and action) are sufficiently “up to” the agent to count as responsible agency, although eliminativist hard determinists doubt mental states have *any* causal powers; see Caruso (*Illusion*). Some soft determinists assert that, had prior contingent conditions (volitions, say) been otherwise, the agent could have done otherwise; hard determinists insist that conditions are never otherwise, thus that the alleged ability is otiose.

Goodman uses a no-self premise (but not on mereological grounds) in his argument against free will, since the 1980s Siderits has been attempting an increasingly nuanced articulation of the implications between the concepts of free will, determinism, and the self,<sup>5</sup> by reference to the theory of “Buddhist Reductionism” that he has extrapolated from the early Buddhist doctrine of “two truths” (conventional and ultimate), a bifurcation paralleled in Western philosophy of science. In this article, I will explore Siderits’s account. Much of the analysis may be applied to Goodman’s account.

### Siderits: Paleo-compatibilism

Siderits’s training is in both Asian philosophy and Western analytic philosophy, but his scholarly efforts are mostly focused on Buddhism. Particularly, he has articulated an impressive, highly complex Buddhist account of the person in his treatise, *Personal Identity and Buddhist Philosophy: Empty Persons*, some implications of which bear on free will. His extremely nuanced, well-argued account involves analyses of the relevant metaphysical, epistemological, semantic, and ethical features of earlier and later Buddhist thought regarding the person, in dialectical exchange with all the relevant arguments and trends in current Western analytical philosophical thought in these fields (*Persons*). Its impressive, complex theoretical structure provides the framework within which his thinking on free will must be situated to be fully grasped.

Over three decades Siderits has presented and published several papers on free will,<sup>6</sup> and his proffered theory has evolved through in-

---

<sup>5</sup> Siderits’s work is classified as middle-period work because his first relevant article appeared in 1987 and, though his most recent paper presentation was at a 2011 conference, most of his relevant published work occurs between the early and recent periods.

<sup>6</sup> See Siderits (“Beyond”; “Buddhism”; “Buddhist”; “Reductionism”; “Expressible”; Panel).

creasingly polished iterations, through the course of which he has come to call it “paleo-compatibilism.” I have struggled to grasp paleo-compatibilism as it has evolved through its iterations, but upon examination of his other relevant works, particularly *Persons*, the richness, complexity, and coherence of his model have become abundantly clear, any limitations on my own comprehension and objections notwithstanding. Many of the questions and objections I raise in the course of my analysis reflect more the stages of my own efforts at understanding than any weaknesses of his account. However, raising them and articulating what a paleo-compatibilist can say in response ought to reveal the explanatory purchase of that theory, and thus ought to help fill in the details of its meaning that might not be apparent in his writings that are more narrowly focused on free will.

It would be misleading to describe paleo-compatibilism as what Siderits actually thinks. He has made it clear that paleo-compatibilism is just a possible view that he has extrapolated from Ābhidharma (that an Ābhidharmika or other sympathetic Buddhist or other philosopher could hold),<sup>7</sup> a view he believes is worth investigating because it evidences the philosophical potential of the Buddhist tradition.<sup>8</sup> In this regard, he may be described as the philosophical analogue of the *amicus curiae*, insofar as he is a friend of the theory, not its actual advocate—an *amicus theoria*. For

---

<sup>7</sup> The Ābhidharma (“higher *Dharma*” or teaching) is the third of three collections of early Buddhist texts (*Tripitaka*, “three baskets”), written in Pāli, that constitute the early Buddhist Pāli Canon; an Ābhidharmika is a follower of the Ābhidharma. The Ābhidharma is a philosophical articulation of the doctrines implicit in the other two sets of texts, the Sūtras (sayings of the Buddha) and the Vinaya (the monastic code). “*Dharma*” (Sanskrit; Pāli: “*Dhamma*”) is difficult to translate, and has many differences in meaning based on usage, but may mean any of the following, loosely: the way things are, the teachings of the Buddha(s), the universal pattern or way, the truth or the path to its realization, and so on.

<sup>8</sup> Siderits (Panel). Siderits takes this general posture—of offering Buddhist ideas as items of potential interest to Western philosophers, without asserting them himself—not only in his articles on free will, but in his major monographs, such as *Persons* and *Philosophy*. His writings on free will, therefore, must be understood within his larger *curriculum vita*, using that phrase in the broadest sense.

ease of reference, however, I sometimes say “Siderits thinks” (and the like) when referring to the implications or claims a paleo-compatibilist makes or would make, but this is to be taken as shorthand for something like “Siderits thinks a paleo-compatibilist thinks,” or some such modified notion.

Siderits admits that his view is not necessarily “the” view Buddhists have articulated on the subject, and he has stated that the doctrine of Buddhist Reductionism that supports paleo-compatibilism is “the view of persons systematically worked out in the Ābhidharma schools” (*Persons* 14). Even in *Persons*, Siderits does not affirm or assert the Ābhidharma Reductionist view, but merely articulates and analyzes it, alongside later Buddhist views as well as recent Western analytic views, the dialectical progression through which leads Siderits to his rich Buddhist account of persons. It is a complex, multistaged view that includes elements of both Ābhidharma Reductionism and Mahāyāna antirealism, as well as what he calls “semantic non-dualism,” an enlightened analogue of common-sense realism that synthesizes the insights of Reductionism and antirealism, but avoids their pitfalls. Siderits’s analysis may be sketched in its progression through four stages. Simplifying greatly, these stages are:

1. *Common sense realism*: perceptual wholes (chariots, lectures, persons, etc.) are real.
2. *Reductionism*: macro-level wholes (conventional reality’s constituents) are ontologically empty, illusions based on aggregations of their more ultimately real micro-constituents, which are not empty. Still, there are pragmatic reasons for speaking as if conventional wholes exist, to the extent that facts at the ultimate level ground conventional discourse (*semantic dualism*).



3. *Antirealism*: even micro-constituents are ontologically empty, thus all is unreal.
4. *Semantic nondualism*: because everything is empty there is no ontological distinction between wholes and micro-constituents, so all is equally real/unreal.

Thus, while paleo-compatibilism extrapolates from the Buddhist Reductionism of the Ābhidharma (stage 2), it does not reflect the full progression Siderits articulates in *Persons*, which incorporates Mahāyāna antirealism and concludes with semantic nondualism. That paleo-compatibilism relies on only an early stage in Siderits's more comprehensive account suggests that a more evolved Buddhist position on free will may be extracted from that larger account, and that paleo-compatibilism is, as Siderits suggests, just an illustration of the philosophical potential of Buddhist thought. Siderits has yet to articulate that more evolved Buddhist view regarding free will, and though it would be interesting to attempt that articulation here, we have enough on our plate with paleo-compatibilism.

Thus, let us explore paleo-compatibilism, first by way of simplifications, and then by increasingly pulling at its problematic dimensions. The main idea in paleo-compatibilism is that the Buddhist bifurcation of reality-discourse into *ultimate* and *conventional* levels, or "two truths," may be used to resolve the tension between determinism and free will. Simplifying, for the Ābhidharmika, *ultimate truth* may be described as reality as it is *independently of conceptual constructions* (although it is perceived by the enlightened mind), and *conventional truth* as whatever we normally take reality to be constituted by, relative to our interests.<sup>9</sup> This view is intimately connected to the mereological reductionism of the

---

<sup>9</sup> Siderits ("Expressible"). In Sanskrit, "*satya*" (or "*sat*") seems sometimes to mean "reality" instead of "truth." Although reality and truth differ, truth-statements are convertible into reality-statements, and vice versa.

Ābhidharma: partite wholes are ultimately unreal, illusions generated by misapprehension or faulty conceptualization of their aggregated, ultimately real impartite parts. To use Siderits's latest example,<sup>10</sup> if only quanta constitute ultimate reality, then in our final ontology there are no lecterns to be found, but because lecterns have pragmatic purposes lecterns are conventionally real even though ultimately unreal. There is nothing in a lectern aside from its aggregated quantum constituents, though those constituents are ultimately real. Philosophers of science might say quanta are "natural kinds," because they identify mind-independent (or, on the similar Buddhist distinction, conceptualization-independent) features of *nature* (the world as it is) carved at its joints.<sup>11</sup>

How does this two truths strategy apply to the free will and determinism issue? Paleo-compatibilism parses dependent origination (putatively, determinism) as an ultimate-level doctrine that asserts the dependence of all conditioned/composite phenomena on previous or simultaneous impartite micro-phenomena (like quanta), and free will as a conventional-level doctrine that asserts agent-based abilities in the pragmatic realm of macro-level objects (like lecterns). The paleo-compatibilist's solution is that determinism and free will are both real, but treated on different reality-discourse levels, or, in a sense, real in different ways, just as lecterns are pragmatically real and quanta are ultimately real. Because realities are necessarily compatible, there cannot be any issue of the incompatibility of free will and determinism, but because they involve different levels of reality, or universes of discourse, to even ask if they are compatible in the usual sense is to conflate discourse domains: the question is semantically ill-formed.

We have, then, an initial understanding of paleo-compatibilism. Now, let us begin to problematize this understanding by examining the

---

<sup>10</sup> Siderits (Panel).

<sup>11</sup> See Bird and Tobin ("Kinds").

notion of compatible realities. Consider the notion of counterfactually possible worlds, worlds closely resembling the actual/factual world that *could have* been a certain way in one respect, *had* the actual world been, say, slightly different in some other respect. For example, imagine a world just like ours in every other respect but which contains a single quantum more than our world. Various counterfactual and other *possible* worlds—realities—may be *mostly* “compatible” with the *actual* world, but any nonidentical worlds,  $W_x$  and  $W_y$ , *necessarily* contain facts that *cannot* obtain in both worlds, or else they would be identical, so any conjunction of all claims about both worlds will contain at least one contradiction reflecting the difference between them. Thus, no nonidentical factual and counterfactual full descriptions of reality are *fully* compatible, though they may be *relevantly* or *mostly* compatible.  $W_x$  and  $W_y$  might be “compatible” in the sense that  $W_x$  contains, say, one extra quantum than  $W_y$  which makes no *significant* difference to anything *else* between them (although both physicists and Mahāyāna interdependence theorists might dispute the qualifier “significant,” given their shared thesis of the interdependence of everything on everything, which implies any difference anywhere will have implications for everything else everywhere). Nonetheless, the conjunction of claims that exhaustively describes nonidentical  $W_x$  and  $W_y$  is inconsistent: it contains at least one contradiction (about the number of quanta).<sup>12</sup>

Philosophers want to know whether both claims—*persons sometimes exhibit free will*, and *the universe is deterministic*—can be simultaneously true in the same way claims about the number of quanta (in  $W_x$  and  $W_y$ ) cannot: on the same bivalent truth scale. Of course, the world-qualified claims “the number of quanta *in*  $W_x$  is  $n-1$ ” and “the number of quanta *in*  $W_y$  is  $n$ ” are compatible, because they are about two different

---

<sup>12</sup> An alternative interpretation that avoids contradiction is to shift from ontology to epistemology, asserting two descriptions of one reality, but some interpretations of the two truths are pointedly ontological (Cozort and Preston “Buddhist” 54-55).

worlds, but the unqualified claims “the number of quanta *that exist* is  $n-1$ ” and “the number of quanta *that exist* is  $n$ ” are inconsistent, as these claims are not indexed, relativized, or partitioned by being about different possible worlds. (The actual world is, *a fortiori*, possible.)

This difference in meaning between relativized and nonrelativized claims is analogous to the difference in meaning regarding the ethical subjectivist claims “ $x$  is immoral *to John*” and “ $x$  is moral *to Jane*” as opposed to the ethical realist claims “ $x$  is immoral” and “ $x$  is moral.” In both cases the indexed, relativized claim-pairs are consistent but the nonindexed, nonrelativized, or absolute claim-pairs are inconsistent. The problematic implication for paleo-compatibilism is that even if the indexed, relativized claims “the universe is deterministic *in ultimate reality*” and “persons sometimes exhibit free will *in conventional reality*” are consistent, the nonindexed, nonrelativized, or absolute claims “the universe is deterministic” and “persons sometimes exhibit free will” remain inconsistent.

Arguably, if conventional reality is *fully* compatible with ultimate reality, then *all* macro-level facts about lecterns and such *are* traceable to ultimate-level micro-facts. But if *some* conventional facts *are not* traceable to ultimate-level facts, then the two levels do not seem *fully* compatible. The paleo-compatibilist can say that conventional items that do not reduce to ultimate items are not conventional *truths*, properly speaking. Thus, the paleo-compatibilist may set aside false conventional claims, and insist that reality admits of two correct levels of description, one of which is not technically true from the perspective of the other, but which remains valid relative to a nonalethic (not truth theoretic) standard—it fits with conventionally recognized norms of linguistic practice. This is, basically, what paleo-compatibilism asserts.<sup>13</sup> But if pragmatically

---

<sup>13</sup> Whereas many Western reductionists treat the reduced level as ontologically inferior, most Buddhist Reductionists treat both levels as *almost* equally real. Some Buddhists

valid descriptions are technically false, we do not seem to have “two truths,” two correct descriptions of reality. Because the traditional compatibilism question is concerned with the consistency of unqualified, nonrelativized claims, that is, with both claims being *equally true*, both traditional compatibilists and incompatibilists alike are likely to withhold assent here.

The idea that some kinds of truths are false is puzzling, although we will discuss some models below in which it is not. It might be better, however, to say—just as Siderits says about the traditional compatibilism question being semantically ill-formed—that the claim “conventional truths are ultimately false” is itself an ill-formed claim (that, perhaps, has only didactic conventional value). This line of reasoning seems to come into play later on in the more developed version of Siderits’s account, when he claims that the two discourse domains are semantically insulated (“Reductionism”), as we shall see. But there are more substantive issues here.

The problem with reducing free will is that conventional things that putatively make up free will—say, deliberations and choices—may be traceable to (thus “compatible” with) neural firings, but if *nonconscious electrochemical determinants* completely account for choices, then choices are causally impotent, unfree. Because the traditional compatibilism problem is *precisely* the worry that nonagential forms of determinism (such as neural determinism) undermine free will, what may be described as paleo-compatibilism’s “compatibility of bifurcated descriptions” seems not only to circumvent the problem, but to obscure and ignore it.

I said earlier we would discuss other models of different truth levels that are less problematic. Thus, Siderits (“Expressible”) makes an

---

consider conventional reality delusional, however, and some consider both identical. See Thakchoe (“Theory”).

analogy between the notion of “fictional truth” and conventional truth, where the former refers to the idea that certain propositions are “true” *within* a work of fiction, such as, say, the lines uttered by Hamlet. Siderits’s analysis appeals to the notion of counterfactuals, whereby, *had* the antecedents (of certain counterfactuals) *been* true, then their consequents (which otherwise involve fictional truths) *would be* nonfictionally true. Simplifying, *had* the play (*Hamlet*) *been* true, then fictional claims about Hamlet *would be* nonfictionally true. By analogy, *if* persons *were* ultimately real, then their choices *would be* free in the responsibility-entailing sense. The self is ultimately unreal, but choices remain free in the responsibility-entailing sense *within the context* of person-discourse, just as although Hamlet is ultimately unreal, he remains highly self-conscious *within the context* of *Hamlet*-discourse. Different discourse domains admit of different conventions governing valid modes of expression and claims.

This counterfactual analysis is promising; however, as argued above (regarding actual versus counterfactual worlds), *counterfactuals* are, technically, *definitionally* incompatible with *facts* in the unqualified or nonrelativized sense. This counterfactual explication is thus vulnerable to the objection that although the counterfactually indexed, relativized claims about agency are consistent with determinism, the nonindexed, nonrelativized claims remain absolutely inconsistent. It also ignores the analogical implication that choices would *only* be free in the responsibility-entailing sense *if* the self *were* real, just as facts about Hamlet would *only* be true *if Hamlet were* true, neither antecedent of which is satisfied.

This also resembles the traditional compatibilist’s so-called “conditional analysis”: an agent is free *if* she satisfies the condition that she *could have* done otherwise *had she wanted to* (to which the critic replies: she never could have wanted otherwise than she was determined to

want). It also suggests that the conventional world is fictional—hardly a basis for concluding free will *exists*. Buddhism does, however, speak of *all* conventional reality as dreamlike: “Buddhahood” is, by definition, *awakening*. Is conventional reality, then, a kind of dream reality? When I fly in dreams, it feels great, but I’m not really flying. Is freedom so? An ordinary person asks an enlightened person: When we awaken, will we have free will? The enlightened person replies: No, but until then you dream you do, and it is appropriate to speak as if you do. The principle of charitable interpretation suggests that this is not what paleo-compatibilism means. The paleo-compatibilist means to convey that the level of description that includes responsible agents is valid, but not as valid as that regarding ultimate reality’s micro-constituents.

Some of the hesitance on the part of traditional compatibilists and incompatibilists regarding the compatibility of items spanning bifurcated discourse domains relates to the complexity of the notion of reduction itself. Thus, let us unpack this notion relative to its contrast terms. In *Persons* (chapter one), Siderits argues that there are three major metaphysical options available for the issue of whether something exists: nonreductive realism, eliminativism, and reductionism—in short, *yes*, *no*, and *sort of*. Nonreductive realists claim that entities of a certain sort really exist as such: “I do believe in ghosts!” Eliminativists claim that certain entities do not exist at all: “There is no Santa!” Reductionists claim that entities of a certain sort turn out on analysis to be entities of another, more ontologically primitive sort: “The genome explains heredity, but it is just a complex pattern of ultimately quantum phenomena.”

This example of reduction is not so troubling. But consider another example that is: “There is a Santa, sort of, but he is just some guy on the department store payroll in a Santa costume, and parents bring their children to sit on his lap while he plays the role of the magical Santa, somewhat like an actor in a play.” Growing up with the belief in San-

ta, only to see it reduced, so to speak, to *this*, provides little comfort—little salvaging of what was thought to be of value. Realism about our minds and souls seems unrealistic, ironically, in light of much of what we know (like knowing there is no real Santa), so it is hard to maintain. Eliminativism is easier to maintain in that regard, but threatens to completely jettison all this is of value in the humanistic conception. Reductionism tries to salvage some of it, but whereas it may salvage the genome and the like, it threatens to cheapen the bulk of what is of value to us, by converting the equivalent of the magical Santa in all that is humanistic into the equivalent of the department store Santa. As the most complicated of the three options, reductionism needs further unpacking.

The nonbasic or “special” sciences may be reducible to quantum terms (hereafter, “quantumese”), but geneticists do not speak quantumese because physicists have not mapped out the genome in quantumese; even if they had, those indefinitely long quantumese conjunctions would be useless to geneticists. Thus, reductionists do not *eliminate* genetic semantics, as doing so would undermine genetic research. Genetic reductionism acknowledges the validity of genetics and the greater validity of quantum physics: genome-talk identifies certain complex patterns of quantum phenomena associated with heredity, which pattern-level is of interest to geneticists. Thus, genetic reductionism represents a middle-path position between nonreductive genetic realism (“genes really exist”) and genetic eliminativism (“genes do not exist at all”). This much seems quite sensible and promising.

However, the conjunction of genome-involving and quantumese reality-descriptions involves no contradictions, unlike the conjunction of claims about  $W_x$  and  $W_y$ . Which of these conjunctions is analogous to the conjunction of claims about free will and determinism? Conventional descriptions of agency depict macro-level behavior (agents deliberating, making choices, acting). To describe the action, “I sit voluntarily on a



chair,” in quantumese would require a computer program few humans could understand, even if nothing was lost, technically, in translation. But to treat people as “just” complex patterns of quanta, however true that may be in quantumese, seems to mischaracterize us. A paleo-compatibilist could reply with an analogy: only one who cannot read sheet music cannot equally well grasp the musical beauty of a Mahler composition on both the level of hearing its performance and reading its notations. The implication is that for the quantumese cognoscenti, however few there may be, nothing is lost in translation, and perhaps precisely because so few of us read the sheet music of ultimate reality (analogically speaking), this justifies the pragmatic discourse domain of conventional reality.

There may be no problem in general here, but we come directly to confront the main problem insofar as Buddhism classically denies that “I” reduces to ultimate-level vocabulary (say, “ultimateese”), which suggests failure of reduction—conventional *falsehood*. It is not even obvious that “voluntarily” reduces; but if something does not reduce, this seems more like the incompatibility of  $W_x$  and  $W_y$  than the compatibility of genes and quanta. The problem is, metaphorically, that the magical Santa does not really reduce to the department store Santa: although the make-believe magical being does not exist (eliminativism), because children see department store Santas, they falsely believe he does exist (error theory). The “I” who is believed to possess autonomy is the equivalent of the magical Santa: a Cartesian soul-stuff-constituted homunculus thought to be entirely nonphysical and to exert telekinetic-like mental powers over the physical brain and body. That entity, like the magical Santa, appears to be *eliminated* in Buddhism (although this is a contentious interpretation that will be discussed further), not *reduced* to some naturalized department store substitute found in its place, the analysis of which explains the former false belief. The autonomy that is conven-

tionally construed as essential to humanism rests significantly on this sort of misconception.

Because “autonomy” conjures up an inflated Cartesian conception of persons most philosophers have outgrown, most writing on free will recently have restricted themselves to the notion of free will “in the responsibility-entailing sense,” or “responsible agency,” for short. The justification for discussing the allegedly anachronistic saw of free will is that most normative institutions and the network of reactive attitudes (resentment, remorse, and so on) presuppose attributions of responsible agency. It seems intuitive that “responsible” will be even more difficult to reduce to quantumese, if not impossible. But free will *may* also be not responsibility-entailing *and* purely physical. Consider the following philosophical fiction or “phi-fi” case (compare “sci-fi”).

Suppose genetically engineered cyborg soldiers, designed to lack hedonic sensation and the left pre-frontal cortical development associated with empathy, are incapable of moral reasoning or Skinnerian conditioning.<sup>14</sup> Suppose these sentient beings can access built-in random-photon-emitting-type neural subroutines to disengage from deterministic programming (making them unpredictable), toggle sensorimotor/neural subsystems on-line/off-line, and so on. It is theoretically possible that such beings are superior to us in terms of self-regulation—autonomy-wise. Thus, there may be free will *not* in the responsibility-entailing sense. Interestingly, cyborg-autonomy, like the genome, presumably *would* correspond to quantumese facts. To the extent Buddhist Reductionism can treat responsible agency as similar to phi-fi cyborg agency, which is presumably reducible, but also account for moral reasoning, say, as a deliberative system’s capacity to recognize and respond appropriately to *moral* reasons (which presumably could be identified in

---

<sup>14</sup> Contrary to most Buddhist thought, hedonic sense and sentience are noncoextensive, for certain primitive life forms with sensorimotor abilities experience no pleasure or pain; see Repetti (Response).

naturalistic terms, for example, on a consequentialist model according to which hedonic values are empirically measureable),<sup>15</sup> it seems possible for free will even in the responsibility-entailing sense to be reducible.

Paleo-compatibilism addresses free will “in the responsibility-entailing” sense, affirming that responsible agents exist at the conventional but not the ultimate level, although facts at the ultimate level ground our attributions of the relevant elements of agency that make it true at the conventional level that, say, Jones performed a certain action and is responsible for it. Siderits sets forth a variety of arguments in support of this bifurcation and why Buddhist Reductionists must opt for a strong version of it according to which the two domains are, as I mentioned earlier, “fully semantically insulated,” a view he calls “semantic dualism” (“Reductionism”). His arguments in support of full semantic insulation appear sound, but they rest on a fairly sophisticated analysis of other doxological requirements of Buddhism implicated thereby; they are extremely complex and subtle, and thus they go beyond our concerns.<sup>16</sup> Nonetheless, I present some challenges below to the notion of semantic insulation that may be raised independently of his extended doxological/theoretical defenses.

The plausibility of paleo-compatibilism seems to boil down to what it means for one of two truths to reduce to another, their semantic insulation notwithstanding (for now). The two truths doctrine admits of many interpretations, which suggest different models of reduction. Let us sketch a few, to place the paleo-compatibilist interpretation in context.

---

<sup>15</sup> Semi-compatibilists conceive the possibility of determined responsible agency as the determined ability of an agent to be reason-responsive, particularly with respect to moral reasons. I have described this reason-responsive ability more specifically along Buddhist ethical lines as “*Dharma*-responsiveness”; see Repetti (“Meditation”).

<sup>16</sup> See Siderits (*Persons*) for the full account, but recall that he is not affirming this view, but rather using it as part of a larger dialectical progression.

For the Mahāyāna schools, ultimate truths are the real way that phenomena exist, their emptiness of naturally being the basis of names, or of being a different entity from consciousness, or of truly existing, or of inherently existing. But for the Hīnayāna schools, ultimate truths are certain kinds of things themselves. (Cozort and Preston 55)

Pre-Mahāyāna<sup>17</sup> Buddhists construe “truths” as *things*, which departs from the analysis of “true” as a predicate of propositions; as Cozort and Preston insist, it “refers to the objects themselves” (54), for example, *nirvāṇa* and *dhammas* (elementary micro-constituents of aggregates).<sup>18</sup> Cozort and Preston summarize four main Buddhist schools of thought on the two truths: Vaibhāṣikas construe conventional truths as nonultimate phenomena and ultimate truths as irreducible atoms and phenomena that remain recognizable even when analyzed into parts; Sautrāntikas Following Reasoning construe conventional truths as nonmomentary phenomena and ultimate truths as momentary phenomena; and Cittamātrins, Svātantrika Mādhyamikas, and Prasāṅgika Mādhyamikas all construe conventional truths as all phenomena other than emptinesses and ultimate truths as emptinesses, although Cittamātrins view emptinesses slightly differently and Svātantrikas also construe nonexistent things (that appear to ordinary people as if they exist) as conventional truths (57). Some describe the two truths as interdependent (because all phenomena are empty), thereby avoiding any dichotomy.<sup>19</sup>

---

<sup>17</sup> Cozort and Preston use the Mahāyāna (“Greater Vehicle”) term “Hīnayāna” (“Lesser Vehicle”) to refer to pre-Mahāyāna Buddhism (most early schools of Indian Buddhism, of which Theravāda is the remaining living tradition), but many find this term loaded.

<sup>18</sup> The lower case term “*dhamma*” is Pāli; it is “*dharma*” in Sanskrit. I use “*dhamma*” only to (acoustically) differentiate between it and the upper case term, “*Dharma*,” only because “*Dharma*” is more widely used among Anglophone speakers.

<sup>19</sup> Thakchoe (“Theory”).

Siderits acknowledges that he is deploying the (Ābhidharma-based, atomistic) pre-Mahāyāna perspective (*Persons* 14), but for didactic purposes he also generalizes, makes comparisons between Western bifurcations of reality and the two truths, and makes somewhat casual glosses covering both kinds of bifurcations when doing so is instructive (call this “Siderits’s first gloss”), as do I. All such comparisons and glosses are relevant to assessing paleo-compatibilism, precisely because paleo-compatibilism is supposed to represent the philosophical potential of Buddhist thought with respect to the traditional (Western) problem of free will. To the extent any Buddhist theory of free will rests on premises not shared by most Buddhists it is vulnerable to objections based on shared Buddhist premises, and Siderits’s first gloss is no exception; this point extends to premises not shared by non-Buddhists. Insofar as I wish to avoid spelling out Buddhist doctrinal differences for every claim Siderits or I make, either about his gloss on the two truths or mine, it will occasionally be useful for me to gloss over doctrinal differences, for ease of reference in assessing the paleo-compatibilist’s *general* strategy of applying the two truths to free will. Thus, I propose a pragmatic oversimplification as “my gloss”: Ultimate truth is *emptiness*; conventional truth is *form* (loosely, everything else, the ultimate nature of which is empty).

Let me explain and further justify my gloss. The two truths—phenomena and their emptinesses—arguably summarize all reality. Some Mahāyāna Buddhists consider them identical. This is the Heart Sutra idea that *form is emptiness*, and vice versa.<sup>20</sup> Abstractly, any plausible Mahāyāna explanations that subsume pre-Mahāyāna understandings into its more inclusive theoretical/interpretive structure apply to my gloss, which is already shared implicitly by some pre-Mahāyāna and explicitly by most Mahāyāna schools. In characterizing two of the three marks of existence, *anitya* (impermanence) and *anātman* (lacking a self or

---

<sup>20</sup> See Gyatso (*Heart*) for an explanation of the Heart Sutra.

independent nature), the Buddha implied that the ultimate truth is that everything is empty, for whatever is impermanent or lacks self (independent nature) is empty. The Buddha also spoke of emptiness in connection with the dependently originated nature of the world and the *khandhas* (the five aggregates that constitute experience);<sup>21</sup> for the dependent lacks independent existence or self-nature and is empty.

Further, although Ābhidharmikas count micro-level *dhammas* as ultimately real in one sense, even such *dhammas* remain empty in another sense because they are extremely ephemeral (momentary) and interdependent (two signs of emptiness), and the macro-level perceivable items they appear to form are ontologically empty—nonexistent illusions of aggregation. Thus, ultimate truth or reality may be glossed as emptiness, and conventional reality as form, or everything else that is empty, which includes all sensory/phenomenal experience. As long as nothing in the argument requires a more doctrinally nuanced specification, it would be irrelevant to object that the *gloss* is, in effect, a *gloss*.

It is tempting to think that the enlightened perceive emptiness whereas the unenlightened perceive form. To the enlightened, all form or phenomenal experience is empty, dream-like; to the unenlightened, all form seems substantive.<sup>22</sup> But the enlightened/unenlightened and ultimate/conventional distinctions are not coextensive. “Chariot” and “Nāgasena,”<sup>23</sup> to use classical examples, are conventional language (“conventionalese”) *constructs* that name aggregations of chariot parts

---

<sup>21</sup> See *Samyutta Nikāya* (SN) 22.95 and SN 35.85, respectively. All references to SN and other elements of the Pāli Canon are available online at the Access to Insight website, available free online at [www.accesstosight.com](http://www.accesstosight.com) (accessed February 21, 2012) and also the Pali Canon Online: The Original Words of the Buddha website, available free online at <http://www.palikanon.org/> (accessed March 26, 2012).

<sup>22</sup> There is an exception for the unenlightened Buddhist who has conceptual but not experiential understanding (Thakchoe “Theory”).

<sup>23</sup> The chariot and Nāgasena examples are derived from the classic Buddhist statement of mereological reductionism, the *Milindapañhā* (“Miln”).

and person-series that are reducible to ultimately existing entities. Others, such as “*ātman*” (nonphysical soul) or phlogiston do not name reducible aggregations. Thus, not all conventional items are ontologically equal: some are more empty or false than others—those that do not reduce. Buddhist Reductionism denies the person Nāgasena exists in ultimate reality but acknowledges that “Nāgasena” denotes validly perceived partite ephemeral phenomena that reduce to ultimately real impartite micro-constituents. Conventional *truths* are the phenomena named “Nāgasena” and “chariot,” or, on an equivalent analysis, correct predications about them, not the unenlightened understanding of those names (as substantive entities). Conventional truths are what are picked out by valid (pragmatic) linguistic devices. Instead of referring to a voluminous conjunction of spatiotemporally indexed concatenated wheel parts, processes, and so on, even the enlightened will say “chariot.”

Illustrating his first gloss on this (“Beyond”), Siderits compares the two truths with a similar distinction in philosophy of science between scientific and common sense vocabularies. For example, science uses “mass” and “mean molecular kinetic energy,” but common sense uses “weight” and “heat.”<sup>24</sup> However, Dennett (*Brainstorms*) identifies a multiple-truth-levels spectrum involving perspectival explanatory “stances,” for example: *intentional* (folk psychology), *design* (algorithmic programming), *functional* (electronic circuits), and *physical* (quantumese). The justification is that level-appropriate vocabularies identify causal phenomena relative to our cognitive apparatus and interests: inherited traits, inflation, and mouse traps are better explained in genetic, economic, and mechanical terms than in quantumese.<sup>25</sup> Siderits himself offers a sustained argument (*Persons* chapter four) that glosses any and all valid nonultimate levels as conventional (call this gloss, which virtually

---

<sup>24</sup> Sellars (“Philosophy”).

<sup>25</sup> Schiffer (“Laws”; “Physicalism”).

amounts to my gloss, “Siderits’s second gloss”). Let us examine Siderits’s second gloss.

Buddhist and scientific reductionism converge and diverge on both levels. Both accept the pragmatic justification of the conventional, yet ontologically privilege micro-phenomena, but whereas Ābhidharma *dhammas* are the Buddhist equivalent of quanta, science would not consider an atom-like (impartite) instance of whiteness, to use Siderits’s example (“Reductionism”), to be a natural kind or quantumese (ultimate). Conversely, Buddhist Reductionists will not likely accept what scientific reductionists consider the ultimate reality constituents.<sup>26</sup> A problem with Siderits’s second gloss is that Siderits argues that for the paleo-compatibilist, “determinism” is ultimatese and “free will” is conventionalese, but the idea that “determinism” is ultimatese is problematic.<sup>27</sup> *Determinism* may be expressible in quantumese, but because quantum mechanics is indeterministic, there is a serious difficulty on the Western side of the comparison. Besides, determinism is an agglomerative abstraction from universal laws that are themselves counterfactual-supporting generalizations about countless partite phenomena, none of which exist in ultimate reality for paleo-compatibilist mereological reductionists, so there is a serious difficulty on the Buddhist side of the comparison. Further, *dependent origination* is centrally about causal relations between *perceptions, intentions, actions*, and the like—all *prima facie*

<sup>26</sup> Even a single (divisible) hydrogen molecule lasts too long to be an Ābhidharma *dhamma*, and quanta are too small. Thus, on my gloss, Buddhist Reductionists who accept physics must conclude that no conventional things (items of valid perceptual/phenomenal experience) have inherent natures that ground the use of names—there are no natural kinds—unless (like atomists who now say physics’ “atoms” are not really atoms) they are willing to say Ābhidharma “*dhammas*” are not really *dhammas*, but physics’ quanta are.

<sup>27</sup> Siderits is espousing the pre-Mahāyāna view, but some Mahāyānists, for example, Garfield, also consider dependent origination ultimate; see, for example, Nāgārjuna (*Mūlamadhyamakakārikā* 24.18), although Garfield’s interpretation of this passage has been seriously challenged, and the more standard Mahāyānist view is that everything in form that is effable—including dependent origination—is conventional. (I owe this observation to Siderits.)



macro-phenomena. Thus, if either my gloss holds, or if we press *quantumese* reductionism on Siderits's first or second gloss (because *dhammas* are not real quanta, they are nonultimate), then the entities posited in dependent origination theory are all conventional. If so, determinism and free will are same-level—conventional— notions, but then there is no warrant for two-tiered theories like paleo-compatibilism.

The paleo-compatibilist, however, could reply that there are degrees along the conventional/ultimate spectrum, or a variety of conventional levels, particularly on the didactic glosses discussed above (the scientific element of which explicitly contains a spectrum or tiered model), and claim that it is confusion among the nodes along this spectrum that accounts for these problematic intuitions, not Buddhist Reductionism. This view, the paleo-compatibilist can claim, is internally consistent, even if Buddhist folk physics, so to speak, as it stands, is inconsistent with quantum physics. Again, the model itself is being extrapolated from *Ābhidharma* simply for didactic philosophical purposes, not for purposes of armchair micro-physics.

What, then, about the free will side of the equation? The paleo-compatibilist thinks “free will” is conventional because it is person-adjectival and persons are convention-level beings, but this does not mean that free will does not reduce: Siderits *does* think it reduces.<sup>28</sup> If the natural abilities that go into free will *do* reduce to *dhammas*, then there seems to be no reason to think free will is an illusion even if we—persons—do not exactly reduce. This is similar to the claim that even if persons are not ultimately real, pains are, despite the idea that pains require subjects to experience them. (Being the subject of experience does

---

<sup>28</sup> Siderits (personal communication, February 2012). If free will were construed as *whole-person-level* action-origination and thus *did not* reduce to ultimatese, however, because, say, there are illusions about the whole-person-level composite-type entity (analogous to illusions about fire's nature as phlogiston), then free will would involve *some* conventional element incompatible with ultimate truth, reminiscent of  $W_x$  and  $W_y$ .

not equate with being a self: enlightened beings lack even a false sense of self, but may experience pain.)

As its meaning unfolds, paleo-compatibilism increasingly resembles traditional compatibilism, which identifies features of action that figure in attributions of responsible agency in beings nonetheless completely governed by nonagential deterministic forces. In later iterations, in which Siderits both coins the label for his model, “paleo-compatibilism,” and sharpens its meaning (“Buddhism” and “Reductionism”), Siderits focuses on the claim that the levels are isolated from, but compatible with, each other. Incompatibilists, however, might object that this resembles calling the tiger and hare “compatible” because they get along *when kept apart in cages*—“cage compatibilism.” The compatibility criterion in the Western discussion is logical consistency (noncontradictoriness), and paleo-compatibilism seems to avoid bivalence the way subjectivism does (by embedding, relativizing, or indexing its claims). But the free will that paleo-compatibilism asserts for conventionalese is libertarian—which is by definition *indeterministic*—and thus the only way determinism and indeterminism *seem* compatible is by being embedded within opaque partitions that obscure their mutual exclusivity.

This resembles a contradiction embedded within a belief system, such as when it is true that Jimmy believes that some proposition is true under one description but believes that it is false under another description: It is true that Jimmy believes that Clark cannot fly and Jimmy believes that Superman can fly. As Michael Barnhart noted, Asian philosophers sometimes exhibit a “blissful maintenance of contradiction,” so we cannot assume that anyone maintaining two beliefs has a *belief about their compatibility*.<sup>29</sup> I have dubbed the violation of this insight “Barnhart’s fallacy,” and the sort of “compatibilism” inferred simply because two beliefs obtain in the same belief system “Barnhart compatibilism”

---

<sup>29</sup> Barnhart (conversation).

(“Earlier”). I do not think paleo-compatibilism is Barnhart compatibilism, but because paleo-compatibilism bears some resemblance to Barnhart compatibilism, compatibilists and incompatibilists alike might resist it.

For such skeptics, although Siderits claims paleo-compatibilism is just an interesting model, he presents it as a viable one, and it is its viability that philosophers who are not paleo-compatibilists wish to measure. Relative to the recent state of dialectic among those philosophers who are not paleo-compatibilists, however, Siderits cannot simply rest one questionable doctrine (paleo-compatibilism) on another (the two truths). In his writings on free will Siderits does not really support the two truths, but in *Persons* he does—in depth and with rigor. *Within* Buddhism it goes without saying that, although Buddhists dispute its interpretation, the two truths doctrine is supported. But Siderits is not only addressing Buddhists; he is specifically modeling the philosophical potential of Buddhism for non-Buddhist philosophers. Anyone who would dismiss paleo-compatibilism should examine Siderits’s *Persons*, which sets his interpretation of the two truths on solid ground. However, his arguments about free will are presented as self-contained journal articles, so although they may be best understood within the framework informed by his *Persons*, they may nonetheless also be assessed as they appear in the self-contained writings in which he has presented them. Let us continue with that assessment, then, by looking at what Siderits has to say about the self.

Siderits explains that Buddhism adheres to an ultimate no-self doctrine that implies that the person lacks an enduring, unchanging nature, and thus is not the same entity at any two instants, so the later stage of this “person-series” does not *seem* to *deserve* the karma from earlier stages (“Beyond” 151). Siderits says it makes pragmatic (conventional) sense to say that all the stages are of the same person, and thus that

the infant and adult Mark Siderits are the same person, but there is no ultimate entity that undergoes the changes, only an appropriate causal continuity connecting them.<sup>30</sup> Thus, he adds, the fact that karmic effects accrue to later stages in the person-series as a function of earlier same-person-series actions poses no injustice, as it might if they accrued to a different person-series.<sup>31</sup> The two truths suggest that to say “I did it and I deserve to be punished” is pragmatically justified (152) and conventionally true—but ultimately (not exactly false, but rather) meaningless. For, as Siderits has remarked, a “statement cannot be true or false if it lacks meaning—in that case it’s just word-salad, not really a statement at all. That’s what happens to ‘I deserve punishment’ when assessed at the ultimate level.”<sup>32</sup> Applying this strategy to the issue of free will, Siderits states:

When we eliminate conventional truth in favor of ultimate truth, we lose our concept of personal identity; but the relations among physical and mental constituents of person-series that ground our use of the concept ‘same person’ remain. I shall argue that we must likewise lose our concept of freedom. (“Beyond” 153)

This is a promising analogy. Thus, if we substitute “free will” for “person” here, we can make a parallel argument:

When we eliminate conventional truth in favor of ultimate truth, we lose our concept of free will; but the rela-

---

<sup>30</sup> Here is just one instance of my earlier claim to the effect that causation plays too heavy a role in Buddhism for too light a conception of causation.

<sup>31</sup> One may object that if there is no identity, nothing can prevent karma from person-series-Siderits from accruing to person-series-Repetti, and there could be no basis for claiming such transference was unfair. The reply would be that such transferences just do not naturally occur, just as pouring water on a basketball just does not ignite it. Arguably, however, only robust causation can prevent such deviant karmic or causal re-distributions.

<sup>32</sup> Siderits (personal communication, February 2012).

tions among physical and mental constituents of volition-deliberation-action series that ground our use of the concept 'responsible agent' remain.

As far as I can tell, this would satisfy compatibilists who thought it close enough to free will in the moral-responsibility-entailing sense.

But why must we "eliminate" conventional truth, if, as Siderits argues (*Persons*), there are good reasons to favor reductionism over the two extremes of nonreductive realism and eliminative nonrealism? Person-series are *also* phenomena, so on my (and possibly his second) gloss they are conventional. The idea that we need to *eliminate* conventional truth reflects the difference between unenlightened and enlightened conventional discourse: we *want* to eliminate *unenlightened* conventional falsehoods (say, about Nāgasena being *ātman*), but we *do not* want to eliminate *enlightened* conventional truths (about "Nāgasena" naming an ultimately impersonal person-series). The two truths—on my gloss, phenomena and emptiness—are both always present and necessarily compatible: phenomena are not *eliminated* by emptiness; rather, they *are* what it is that *is empty*. But if paleo-compatibilists want to affirm that *some* conventional truths are *not* reducible to ultimate truths, and thus they are eliminated on reductive analysis, then this is like  $W_x$  and  $W_y$ , but then the two truths *are not really* compatible.

Siderits seems to mean that persons reduce to person-series and responsible agents reduce to responsible-agent-series, so they are not really "eliminated," but rather just understood more clearly for what they are. Perhaps Siderits was speaking causally when he said "eliminate" here (perhaps he meant "substitute by more accurate language" or some such alternative). This casual mode might also explain some of the discrepancies that seemingly emerge when all his writings on free will are treated as a single series of arguments. For in the first iteration of his views on free will ("Beyond"), Siderits argues that we neither *have* nor

lack free will at the ultimate level because there is no ultimate person, and a nonperson cannot be free or unfree (just as a rock can be neither).<sup>33</sup> Thus, Siderits concludes, the free will question *does not arise* at the ultimate level (155).<sup>34</sup> Siderits says Buddhism therefore goes “beyond compatibilism.” Shortly afterwards, in entertaining what an opponent might say, Siderits says: “If ultimately there are no persons but only physical and mental events in a complex causal series, then the ultimate truth about ‘us’ must be that ‘we’ are not free” (“Beyond” 158). But Siderits’s point is that it is *that* sort of reasoning that would conflate the levels. For we cannot say “we” are the person-series (and, given that “our” volition-deliberation-action sub-series are appropriately auto-regulating, “we” are “free”) because “there cannot be identity between ultimately real entities and entities that are only conventionally real. But this is not to say that ‘we’ do not reduce.”<sup>35</sup>

Reduction and identity are not the same. But while this makes sense when something conventional disappears at the ultimate level under mereologically reductive analysis, it does not seem equally intuitive in cases where nothing seems to disappear, for example, when water reduces to H<sub>2</sub>O: Is not water identical to H<sub>2</sub>O? A plausible, negative answer could draw on the distinction between “wide” and “narrow” conceptions

---

<sup>33</sup> Two objections: First, there is a sense in which a causal stream may be said to be “free” if it is unobstructed, say, by a dam. In support of Siderits’s claim, however, Frankfurt (“Freedom”) argues that “free will” applies only to volitional beings that can suffer its privation. By analogy, rocks cannot be *blind*. Second, consider a flawed claim analogous to Siderits’s claim that we neither have nor lack free will because only persons either have it or lack it: we neither experience pain nor do not experience pain because there are no selves (subjects of experience, experiencers) in ultimate reality (as if person-series that are not persons cannot experience pain). In Siderits’s defense, it may be said that because there are no subjects of experience in ultimate reality, it is neither true nor false that such subjects experience pain. But there remains a sense that even if subjects are not selves in the sense denied by Buddhists, there may be pain experiences in ultimate reality.

<sup>34</sup> But on my gloss and perhaps also his, *no questions* arise at the ultimate level, so this *prima facie* plausible solution is somewhat misleading.

<sup>35</sup> Siderits (personal communication, February 2012).

of semantic/psychological content, for example, where, arguably, on the wide conception “water” has a hidden indexical that tags water-like stuff in the proximal (wide) environment that played a causal role in the formation of the water concept, as opposed to some water-like concept that exists entirely in the (narrow, Cartesian) head, to paraphrase Putnam (“Meaning”). Putnam’s ingenious “Twin Earth” scenario depicts a *doppelgänger* with a narrowly identical (psychologically indistinguishable) water concept but a widely different (environmentally distinct) water concept, say, because the (water resembling) substance there (which behaves perceptually identically) is composed of xyz. If the concept is narrowly identical, but widely different, then, arguably, narrow water reduces to wide water, but is not identical—it cannot be, because H<sub>2</sub>O and xyz are not. In this (or a similar) way, then, reduction and identity are not the same.

What of the abilities that constitute free will? Siderits’s analysis of volitional regulation resembles Frankfurt’s, such that free will involves approval of the volitions on which one acts.<sup>36</sup> However, Siderits treats *volition* as if paleo-compatibilism parses it as ultimate, person-series-level stuff, not conventional. Presumably, anything may be analyzed at an ultimate or conventional level, or a mix of both, as per our spectrum and degrees discussion above. But Siderits’s opponent again might charge that if person-series and deliberation-series are both ultimate, then a conventional person that has conventional free will *just is* an ultimate person-series that exhibits ultimate volition-deliberation-action-series regulation. Siderits would reply that this sounds like paleo-compatibilism enough for his purposes, except that it slightly misunderstands the “just is” in the Buddhist Reductionist (semantic dualist) way—

---

<sup>36</sup> On Frankfurt’s model, when an agent acts on a volition the agent approves of, this exhibits an effective meta-volition (a volition about a volition), and free will involves such effective (volition-controlling) volitional/meta-volitional harmony. For a Buddhist account along similar lines see Repetti (“Meditation” and *Counterfactual*).

not as attributing identity, but as asserting certain relations among different kinds of utterance. This is subtle, but it will do.

If person-series and volition-regulating deliberation-series are not ultimate, however, but conventional, then conventional beings have conventional free will. Either way, we *might* have free will without levels-conflation or even levels-crossing. The volition-*regulating* view of free will is arguably conventional, because conventionalese is not *about* emptiness, but *phenomena*, such as “red apple.” On my gloss, perceptual redness is a valid cognitive phenomenon. Thus, the claim “red apples exist” is (valid) conventional.<sup>37</sup> Likewise, on my gloss, volitional functions that constitute free will on the analysis Siderits implicitly accepts (such as approving one’s volition) are conventional because they are not *about* the emptiness of phenomena, but the (introspectively and interoceptively) perceived phenomena.

There seem to be many ways in which something may be understood as conventional, and shifting among these senses may be what generates the appearance of discrepancy. As suggested earlier, it may be reasonable to view conventionality along a spectrum, where, for example, persons, tables, and other macro-level mereological wholes are at the far end of the conventional side, their smaller parts are still conventional (because they themselves are not impartite) but are closer to the ultimate side, and the impartite and the emptiness of everything is at the far end of the ultimate side. In the didactic articulation of many of the ideas in this discourse, both in Siderits’s specific claims and his glosses, as well as in my analyses, claims are made about items occupying different nodes along this spectrum, and that is likely what gives rise to the

---

<sup>37</sup> However, for some (Ābhidharmikas), perceived redness is a *dhamma*, a basic pattern, and thus ultimate, though perceived apples are not *dhammas*, so they are conventional. (This observation illustrates the tedium that would otherwise ensue regarding spelling out all the doctrinal differences for each claim here, absent both my and Siderits’s glosses.)



appearance of discrepancy. Let us keep this in mind as we examine both paleo-compatibilism and possible objections to it.

Thus, suppose a monk, asked why he gives away food while collecting alms, claims, “Because I feel sorry for the destitute.” The monk’s claim describes phenomenal experience with *mudita* (sympathetic appreciation) and its volitional dynamics leading to his actions. The claim is not about the emptiness of the phenomena, but the phenomena, so it is a conventional claim that might satisfy compatibilist criteria and (along with similar criteria satisfactions) render the act conventionally free. Further, if the monk’s reasons are conventional, then those reasons are also reasons to think that there are conventional persons, like the monk, to whom the (conventionalese) *Vinaya* (the Buddha’s monastic code) applies. For the same analysis that supports the idea that conventional beings possess free will implies that such beings are conventional persons: phenomenal beings that possess self-regulating causal powers (for example, anger control)<sup>38</sup> that conventionally impersonal-type series—such as rivers—lack.<sup>39</sup> Siderits’s claim is not that free will involves levels-conflation, but that the traditional compatibilist/incompatibilist debate does, in disputing whether (conventional) free will is compatible with (ultimate) determinism. But a traditional compatibilist, reasoning along the lines depicted in the monk example, could say that the volitional causes that determine choices are on the same level as the person, conventional, so there is no levels-crossing; so could anyone who considers everything but emptiness conventional.

Siderits later sharpens this contrast, asserting “full semantic insulation between the two discourses” (semantic dualism) (“Reduction-

---

<sup>38</sup> Śāntideva (*Bodhicāryavatāra* 6), referenced in the bibliography as Shantideva and Padmakara (*Bodhisattva*), discusses anger control.

<sup>39</sup> Frankfurt (“Freedom”) defines a “person,” roughly, as a being with meta-volitions and nonpersons as beings that lack meta-volitions, such as animals that simply act on whatever volition arises.

ism” 36). But how can one level of reality relate to another if the languages used to describe them cannot make contact? Siderits suggests that conventionalese “supervenes on” ultimatese (“Buddhism”). Though in his writings on free will Siderits does not explicate how this abstract relation *connects* insulated discourses, he does explain what sort of supervenience he has in mind elsewhere (*Persons*, chapter one). He also explains why he thinks it is important to keep both discourse domains semantically insulated, noting that this very insulation poses a problem in terms of being able to express the two truths without violating the insulation (“Expressible”). Siderits’s keen analyses (involving the notion of meta-language, too complex to repeat here) indirectly support the claim that his parsing of the two truths is required for Ābhidharmika Reductionists, but not necessarily other Buddhists.

We cannot summarize all his arguments for Buddhist Reductionism, but his distinction between reductive and nonreductive supervenience is relevant (*Persons*). Reductive supervenience is the idea that the items in the domain to be reduced, D2, to those in the more basic domain, D1, map onto those items in an appropriate way, such that the facts at D1 both determine and explain those at D2. This sort of mapping relation is fairly unproblematic when a branch of science easily reduces to more basic science, such as chemistry to physics, but problematic when mental states are said to supervene on neural states, particularly for those who think the latter cannot adequately explain the former. This is where nonreductive supervenience may be brought in, as the claim of the determination of D2 by D1 (no change in D2 without one in D1), but where everything about D1 is insufficient to explain everything about D2, so D2 has some sort of explanatory autonomy.

To get a handle on this, we might think of things connected with the narrow conception of water as at D2 and those connected with the wide conception of water as at D1. Some philosophers of mind have

claimed, for example, that although some mental states, such as volition, supervene on physical ones in which they are *realized* (say, in the configuration of synaptic firings instantiated synchronically with that volition), these mental states remain unexplained by the physical states because any *intentional* state is always “about” something, whereas purely physical items are never “about” anything: electrochemical neural discharges lack intentional content—meaning—or what Searle called “aspectual shape” (*Mind* chapter 7). This might be a case of nonreductive supervenience, and, as Siderits notes, some early Buddhists possibly held this view of persons.<sup>40</sup>

It is commonly accepted that different neural-state configurations might instantiate that same volition and that other instances of the same neural configuration might lack that volition if an alternate-world-inhabiting (atom-for-atom-identical) *doppelgänger* could lack consciousness altogether. To the extent semantic dualism is akin to mind-body dualism, it shares such mapping problems. Anyone who believes in free will seems implicitly committed to the reality of conscious mental states, and neuroscience places a burden on the folk psychologist to come up with something like a supervenience relation. To Siderits’s credit, he offers plausible arguments for reductive supervenience; but they are incredibly complex and subtle, and this issue is too problematic to resolve here.

Thus, just as it may beg the question (in the “looser” sense that the premise is as problematic as the conclusion, as opposed to question-begging in the tighter sense of circular reasoning), contextually, to appeal to the problematic two truths doctrine to support a position on free will, so too it may beg the question to appeal to supervenience to support semantic dualism. We now have stacked puzzles—free will, two

---

<sup>40</sup> Pudgalavādins (those affirming a “pudgala” or “person” doctrine) may have held this intermediary view (between nonreductive realism and reductionism); see Siderits (*Persons* 89–91).

truths, semantic dualism and supervenience. Such question-stacking appears as a weakness, but in fairness to Siderits it may be—and I think it actually is—indicative of the broadly coherent, multidimensional explanatory purchase of his theory. It ought not to be incumbent on readers of Siderits’s individual articles on free will to digest his entire corpus on Reductionism in order to assess his *amicus* brief on behalf of paleo-compatibilism, and we can only devote so much space to that larger work here. Those who wish to more fully explore this larger project are encouraged to direct their attention to Siderits’s brilliant treatise, *Persons*. Having noted the extent to which paleo-compatibilism is grounded in this larger dialectical and explanatory framework, then, let us focus the remainder of the assessment of paleo-compatibilism on its articulation in Siderits’s articles on paleo-compatibilism.

Siderits says that at the ultimate level “there are events that correspond to what we ordinarily call deliberating and willing” (“Beyond” 155), and this is why he thinks the volitional analysis applies at the ultimate level. Does free-will-supporting conventionalese about volitional regulation map onto ultimatese about micro-level volitional phenomena, and does conventionalese about persons map onto ultimatese about person-series, such that conventionally responsible agents reduce to ultimate series that ground their conventional reality? This seems to be what Siderits intends, and that would be a welcome argument from the compatibilist perspective, but in light of the complexity of everything he says that bears on the subject, it is not entirely clear.

Siderits’s suggestive approach has great merit, but some of that merit is obscured by the complexity of its many details:

To see myself as capable of acting freely, I must view myself as the sort of entity that can endorse the actions it performs. This requires that I take deliberating and will-

ing as abilities or faculties that I possess. This possibility is denied us at the ultimate level of truth. (155)

Siderits thinks “this possibility is denied” because “persons” do not exist ultimately, though we exhibit these abilities. Siderits continues:

At that level there are events that correspond to what we ordinarily call deliberating and willing; but there are no persons who are the authors of those events, for a person is a mere conceptual fiction that is constructed out of those and other events . . . . Thus, it is only at the conventional level that an action may be said to be performed freely. (155)

But if “a person is a mere conceptual fiction”—a mental entity—“that is constructed out of those and other events,” where those events are mental, that is like saying *shoes are not leather because there are no shoes, only leather-shoe parts*.<sup>41</sup>

There seems to be an implicit asymmetry in the treatment of items here if, say, conventional volitions are said *without qualification* to correspond to ultimate events, but persons and free will are *qualified* as being “only” conventional. If there are ultimate events that correspond to conventional volitions that, *arguendo*, might satisfy some criteria for free will, then cannot ultimate events also correspond to conventional free will? If there are events at the ultimate level that correspond to deliberations, then there are arguably events at the ultimate level that correspond to the sorts of things that constitute responsible agency—that correspond to free-will-exhibiting agents (persons). Conversely, the same reason to think there is no ultimate free will or ultimate person

---

<sup>41</sup> Arguably, this objection involves a straw man fallacy, where the intended claim, charitably interpreted, is that *it is not ultimately true that shoes are made of leather because shoes are not ultimately real*. But the objection may be restated more charitably: however empty shoes are ultimately, they are made of leather. The counterexample thus survives.

may be used to conclude that there are no ultimate deliberations, as opposed to, say, deliberations-series. Persons are changing series of psychophysical events, just like deliberations are, so why is there reductive asymmetry between them? Asymmetry—per se—is fine, but not when things seem symmetrical.

The paleo-compatibilist might again appeal to the graded macro/micro, conventional/ultimate spectrum to explain any appearance of asymmetry: persons, person-series, deliberation, deliberation-series, volition, volition-series, and so on, down to *dhammas* and emptiness. Perhaps our shifting focus along this spectrum is what is generating the asymmetry, not the theory. Thus, let us examine this more carefully by focusing on just one centrally relevant item here, deliberation.

Deliberation is minimally dyadic: it minimally requires consideration of two alternatives. Even if superior beings apprehend many alternatives simultaneously, we do not. If each consideration corresponds to ultimate micro-events, the only way this micro-series can count as “deliberation” is if *something* links them together, even if that *something* is just a combinatorial or causal pattern/formula. Similarly, it seems there can be no “mother” if there is nothing that links together the bearer of a zygote and the woman who delivers a baby (*Miln*). But if this *something* does not exist at the micro-level, the claim that deliberation corresponds to ultimate events is problematic. If it does exist at the micro-level, then ultimate reals *form* micro-level deliberation-series that ground the conventional reality of deliberations; however, then ultimate reals could equally *form* micro-level responsible-agency-series that ground the conventional reality of responsible persons. At issue is this *asymmetry* in mereological reduction.

In his original article (“Beyond”), Siderits *implicitly* accepted mereological reductionism; now he *explicitly* asserts that “Buddhist Reductionists are thoroughgoing mereological reductionists” (“Reductionism”

34). This implies that “persons are reducible without remainder to completely impersonal entities because all partite entities are reducible without remainder to their parts” (34). But then the question is: Why are some reductions privileged, but others not? Buddhist mereological reductionism is articulated classically by Nāgasena: chariot parts have no relationship with each other independent of those imposed by our needs, so chariots are not natural kinds (nothing inherent in them grounds the name “chariot”), but pragmatically justified linguistic convention names their configuration “chariot,” as with the label for the psychophysical series “Nāgasena,” the configuration of elements of which lacks self-nature (*Miln*).

Mechanical artifacts intuitively lack any whole-level feature that is irreducible to their parts, but Nāgasena’s intentionality and consciousness seem irreducible to their parts. This apparently compositional difference would at least provide *some* ground for mereological asymmetry, but Buddhist Reductionism denies any such ultimate difference *here*, as Buddhism more generally decomposes even consciousness and the like into the interdependent interaction of sense-organs and objects of experience. If “all partite entities are reducible without remainder to their parts,” however, then just as there is equally no chariot or person, so too there are no Buddhas, volitions, karmic merit, thoughts—eyes reading this—or anything apart from the scientific or Buddhist version of quanta, possibly also excluding quanta if they are divisible even into homogeneous parts. But as with any other *reductio ad absurdum* argument, the allegedly absurd conclusion may be embraced, and Buddhist Reductionists do embrace this conclusion. They dispel its apparent absurdity by saying that all these things that technically do not exist in ultimate reality nonetheless do exist conventionally. This suggests an ironic use of cliché: Buddhist Reductionists apparently can have and eat their (ultimately nonexistent) cake too.

Reductionists claim the person is a mere conceptual fiction, but maybe reductionism is in a similar category. Reduction involves identifying micro-constituents as ultimately real, but those micro-constituents must either be indivisible/impartite entities (atoms)<sup>42</sup> or divisible/partite entities (nonatoms).<sup>43</sup> Both options are deeply problematic.<sup>44</sup> If micro-reality is nonatomistic, then for mereological reductionists all macro- and micro-levels are partite—equally unreal—and this ontologically egalitarian option unveils no ontologically privileged ultimate reality. Thus, mereological reductionists—who affirm that all macro-level partite wholes are ultimately unreal and that only micro-level impartite entities are ultimately real—must reject either mereological reductionism or nonatomism. That leaves recalcitrant mereological reductionists one option, atomism.

Suppose there is an atomic level, where, say, all items are homogeneous quanta. However, as some Tibetan lamas say,<sup>45</sup> even the smallest particles can be distinguished into north, south, east, and west sides. This makes sense because if they have no sides, as Parmenides noted, there cannot be a plurality of them, but Ābhidharma posits a plurality of *dhammas*. (Without anything to differentiate them, no less, it is logically impossible that they could aggregate to even appear as anything that

---

<sup>42</sup> Ābhidharmikas assert that only *dhammas* are ultimately real, but this supports my gloss that *everything else* (all the phenomena of experience, *except* its perceivable *dhamma* fragments, if any) is, by default, conventional.

<sup>43</sup> Mahāyānists deny that *dhammas* are ultimately real, insisting that the ultimate truth is that reality is empty, which also supports my gloss that *everything else* (apart from emptiness) is, by default, conventional. In *Persons*, Siderits seems to favor this conception insofar as it occupies a node along the progression of his four-staged argument closer to its final conclusion.

<sup>44</sup> This approach is intuitively physical, one might object, but Ābhidharmikas are explicitly committed to the existence of nonphysical *dhammas*. Siderits's argument does not assert the existence of nonphysical *dhammas*, but neither does he deny their existence. Siderits makes clear, recall, that he is only extrapolating his model from Ābhidharma, not that any Ābhidharmika actually advocates paleo-compatibilism. Nothing in paleo-compatibilism excludes the possible existence of purely mental (nonphysical) *dhammas*.

<sup>45</sup> I owe this reference to Tibetan lamas to Dan Cozort.



may be differentiated.) Thus, they must have sides, so each such entity admits of regions (parts), each region itself admits of regions, and so on, *ad infinitum*. Thus, arguably, there is no *true atom*, thus no impartite, ontologically superior, ultimate reality. Where does that leave the recalcitrant mereological reductionist?

One way to try to evade this objection would be to say that truly impartite atoms are dimensionless points that have no sides, regions, or parts. That will not work because dimensionless entities are nonphysical, but nothing lacking physical magnitude can aggregate to form anything at the macro-level, and, conversely, nothing that has physical magnitude can be reduced to anything that lacks it. Thus, mereological reductionists affirming ultimate reality must reject atomism along with nonatomism. Thus, mereological reductionists seem required to admit that there is no mereologically ultimate reality.

Reasoning like this supports Mahāyāna antirealism, but perhaps paleo-compatibilism's atomistic "tropes" might circumvent the above logical dilemma, such as instances of whiteness (all of which are homogeneous and thus recognizable even if analyzed into identical segments); the Ābhidharma suggests a large but finite number of them aggregate for a perceivable moment of (white-perceiving) experience. But if so, it is not that *nothing partite/composite is real*, but that *nothing perceivably heterogeneous is real*. A *regular coffee*, once blended, *appears* perceivably homogeneous, thus ultimate, though mereologically divisible into black coffee, milk, and sugar (suppose, *arguendo*, that these constituents are themselves homogeneous), just as green appears perceivably homogeneous, though mereologically divisible into yellow and blue. If Buddhist Reductionism is amenable to scientific discovery, it becomes problematic because photons, photo-receptive optic cells, and so on do not seem hospitable to the homogeneous-color-type model of trope-atoms.

Atomistic reductionism *posits* that reductive divisibility comes to an end with ultimate trope-like atoms of reality, but this posit—science aside—remains undermined by the logical dilemma that anything *with* magnitude is divisible and anything *without* magnitude cannot be aggregated. Because perceivable phenomena have magnitude, there are no atoms; thus, again, for the mereological reductionist there is no ultimate reality.

The Buddhist Reductionist can claim we hit rock bottom when we arrive at trope-quanta, the first micro-level populated by *homogeneous* entities, because further divisions only exist *mathematically/conceptually*, and Buddhist ultimate reality is what exists *independent of our conceptualizations*. However, it seems unprincipled—a distinction without an *ontologically relevant* difference—to *ontologically* privilege homogeneity (black coffee) over heterogeneity (regular coffee), particularly when we cannot *conceive* one without the other. Such a preference seems—*contra* the Buddhist Reductionist—to rest on some sort of pragmatic, psychological, hence conceptualized item that is intuitively nonprivileged in conceptualization-independent reality. Could this be the presupposition that grounds the intuition that ontologically differentiates persons asymmetrically from deliberations and volitions, as the paleo-compatibilist thinks the former “only” conventional and the latter ultimate? But even if this were not problematic, deliberation is dyadic, not homogeneous/monadic; so is volition, because it includes (a) conative/teleological impulse, toward (b) an intentional object (desired object, experience, or state of affairs). If *these* decompose into homogeneous/monadic parts, however, there can be no *deliberation-trope* or *volition-trope* among them, just as there are no *mother-tropes*, so it is a stretch to say “there are events that correspond to what we ordinarily call deliberating and willing” (or mothering) in ultimate reality.

Admittedly, similar difficulties challenge many competing Buddhist and non-Buddhist metaphysics, so some of these difficulties are not peculiar to Buddhist Reductionism. But to the extent paleo-compatibilism rests on Buddhist Reductionism, it faces such difficulties, regardless of how widespread such difficulties may be. Siderits's more comprehensive four-staged Buddhist synthesis, however, likely resolves or circumvents most if not all of them (and actually rests its move from Buddhist Reductionism to Buddhist antirealism on similar considerations), but Siderits has yet to articulate the implications of that account for free will.

Turning to free-will-specific claims presented in Siderits's more refined iterations of paleo-compatibilism, Siderits distinguishes four paleo-compatibilist theses. The first two are libertarian;<sup>46</sup> the last two are hard determinist:

1. Persons are free in the responsibility-entailing sense.
2. Freedom requires that persons be originating causes.
3. Nothing could be an originating cause in the required sense.
4. All psychological states are the effects of prior causes.  
("Buddhist")

Siderits argues that (1) and (2) are conventionally true and entail libertarianism, and (3) and (4) are ultimately true and entail hard determinism ("Buddhism"). The paleo-compatibilist considers ultimate hard determinism compatible with conventional libertarianism because both are true at different levels.

---

<sup>46</sup> Libertarians think free will and determinism are incompatible; agents are free because they are originating causes; and thus determinism is false.

On my gloss, both free will and determinism are on the phenomenal/form level (conventional). A related Mahāyāna interpretation that may be used to articulate a Buddhist soft determinism is *interdependent origination*.<sup>47</sup> Mahāyānists argue that there is no potter without a pot, no action without an agent (action-performer), and thus the agent/action pair is interdependently constituted by the action-performance. This conception provides a less inflated conception of the person, one that seems intuitively not in need of reducing. Let us thus define “*Agents<sup>M</sup>*” (where the superscripted “M” represents “in the Mahāyāna sense”) as such interdependent action-performers. Thus, we can substitute “*Agents<sup>M</sup>*” for “persons” in Siderits’s (1) and (2), yielding deflationary statements consistent with *soft* determinism:

1. *Agents<sup>M</sup>* are free in the responsibility-entailing sense.
2. Freedom requires that *Agents<sup>M</sup>* be originating causes.

As even Theravāda scholars have argued, the Buddha thought volitional actions entail karma (merit) because they are voluntary, despite how strongly shaped by prior karma, and that they are initiating causes, *contra* the fatalists, because they create new karma and make a difference to the event series.<sup>48</sup> Thus, (3) and (4) are arguably false, and we may substitute the implied “persons” in (3) with “*Agents<sup>M</sup>*”:

3. *Agents<sup>M</sup>* can be originating causes in the required sense.

Let us dub “paleo-soft-determinism” the conjunction of these modified theses (1)-(3). Ockham’s razor seems to favor paleo-soft-

---

<sup>47</sup> An authoritative source for the doctrine of interdependence is Nāgārjuna (*Mūlamadhyamakakārikā*), an early-period source is Kasulis (*Zen*), and recent-period sources are Gier and Kjellberg (“Buddhism”) and Wallace (“Buddhist”). Wallace discusses a Mahāyāna model of interdependence, consistent with physics, in which every quantum is holographically interconnected.

<sup>48</sup> See Harvey (“Freedom”) and Federman (“Buddha”).

determinism over paleo-compatibilism,<sup>49</sup> insofar as libertarianism rejects scientific determinism and embraces mysterious causation. The paleo-compatibilist seeks to save the notion of free will implicit in the network of beliefs, judgments, and reactive attitudes that constitute our normative institutions and practices,<sup>50</sup> a fine intention consistent with the Mahāyāna notion (expressed in Siderits's final synthesis as semantic nondualism) that conventional reality is as it appears. I share this goal, but dispense with the unnecessarily heavy metaphysical "baggage," as Fischer puts it (*Way*), of libertarianism.

In his latest relevant publication ("Reduction"), Siderits condenses (1)-(4), and maintains free will is conventional and determinism ultimate:

It can be true both that (1) persons are sometimes the originating causes of their actions, for which they are then responsible; and (2) each of the impermanent, impersonal elements in a causal series of psychophysical elements is causally determined by earlier elements. ("Reductionism" 36)

Let us call these "new-(1)" and "new-(2)," to set them apart from (1)-(4). Referring to new-(1)'s libertarianism and new-(2)'s determinism, Siderits says new-(1) "is conventionally true" and new-(2) "is ultimately true" (36). Because they are true in "fully semantically insulated" discourses, new-(1) and new-(2) "cannot be incompatible" (36)—inviting critics' suspicions of cage compatibilism.

---

<sup>49</sup> Ockham's razor, in Indian philosophy "the principle of lightness," is the principle of explanatory parsimony: always prefer the least risky hypothesis, or, in some interpretations, prefer the one with the fewest ontological commitments. See Baker ("Simplicity").

<sup>50</sup> Siderits (personal communication February 2012).

As with (1)-(3), new-(1) and new-(2) admit of inflationary and deflationary readings. If read in *inflated* terms of *libertarian* agents and *hard* deterministic causation, as Siderits does, new-(1) and new-(2) seem contradictory, as new-(1) is *indeterministic* and new-(2) is *deterministic*. Someone, *S*, might *believe* some proposition, *P*, and also *believe* its negation,  $\sim P$ , if they manage to partition (semantically insulate) these beliefs from each other enough to not notice the conflict, but whereas the belief statement “*S* believes *P*” is consistent with the belief statement “*S* believes  $\sim P$ ,”<sup>51</sup> the simple statement “*P* and  $\sim P$ ” remains contradictory. But, if read in the deflated terms of interdependent *Agents*<sup>M</sup> and actions, new-(1) and new-(2) are compatible in the semantically exposed (un-insulated) sense; then, paleo-compatibilism is otiose.

To support paleo-compatibilism, Siderits quotes the (ironically) Mahāyānist Śāntideva, who presents a series of statements (22-31), only one of which I repeat here (22), instructing Buddhists to see others’ behavior as—according to Siderits—determined by *impersonal* factors, the way bile is, for purposes of anger prevention. Śāntideva then entertains an inconsistency objection (32), to the effect that behavior’s *impersonal* causation contradicts the *personal agency* (arguably, free will) that would be required for anger prevention. In reply (32), Śāntideva differentiates between anger’s impersonality and the necessity of the assumption of personal agency for the Buddhist path.

22. There is no anger in me toward bile and the like  
though they cause great pain.

Why anger toward sentient beings? Their anger is also  
due to causes....

---

<sup>51</sup> Insofar as belief embeds propositions in semantically opaque contexts (propositional attitudes), it makes it possible for contradictory propositions to occur unnoticed; recall Jimmy’s contradictory beliefs about Clark and Superman.

32. [Objection:] Prevention [of anger] is thus not appropriate, for who prevents what?

[Reply:] It [prevention] is taken to be appropriate with regard to dependent origination due to the cessation of suffering. (“Buddhist”; “Reductionism” 31)<sup>52</sup>

Siderits thinks verse 22 supports determinism: we do not get angry at *impersonally caused* things, and others’ behavior is *impersonally caused*. This is *consistent* with *hard* determinism, but not an *argument* for it.<sup>53</sup> It is also consistent with *soft* determinism. Śāntideva might mean that most worldlings<sup>54</sup> are so afflicted by (the three poisons of) greed, hatred, and delusion that they are *virtually* hard determined, which warrants exculpatory reactive attitudes such as nonanger and compassion. Buddhists can change behavior upon *Dharma* reflection, an ability arguably better accounted for by *soft* determinism.<sup>55</sup> The worldling may alter behavior under similar circumstances, so *virtually* hard determined behavior can remain *soft* determined if it involves certain unexercised abilities. No statement consistent with two theories favors one over the other, *ceteris paribus*.

---

<sup>52</sup> Quoting Śāntideva (*Bodhicāryavatāra* 6).

<sup>53</sup> Siderits says he does not think this is an argument for hard determinism, but that what Śāntideva develops in the subsequent verses is an argument for the conclusion that responsibility-entailing freedom cannot exist at the ultimate level; Siderits thinks that hard determinism would follow from this only if you were an eliminativist about persons, whereas the Reductionist is not an eliminativist. Siderits (personal communication, February 2012). But he does *seem* to take paleo-compatibilist theses (3) and (4) above to entail *hard* determinism, so either he does not draw the hard element from Śāntideva or I have misinterpreted him to take (3) and (4) to entail hard determinism. Another way to put this is: if one does not think the determinism in play is hard, then why would one think there’s a need to show it is paleo-compatible?

<sup>54</sup> A “worldling” is a person who has not attained the first level of direct spiritual realization/transformation, that of “stream entry,” typically precipitated by Buddhist meditative discipline. Thus, most non-Buddhists and many (maybe most) Buddhists are worldlings, though non-Buddhists can be stream entrants. I stipulate that I use the term loosely to refer to most worldly folks and/or all who are not stream entrants.

<sup>55</sup> For a defense of this claim, see Repetti (“Meditation” and *Counterfactual*).

Siderits thinks verse 32 supports the paleo-compatibilist claim that ultimasese hard determinism does not conflict with the needs of unenlightened, conventionalese-speaking Buddhist practitioners, who must adopt the conventional notion of efficacious agency to follow the Buddhist path, thereby evidencing *some* authoritative support for paleo-compatibilism. However, if Śāntideva *did* mean that impersonally caused behavior was *hard* determined, then Śāntideva's response to the inconsistency charge arguably cannot be easily dismissed simply by reference to the Buddhist hope of reducing suffering. The idea that *by doing something* (contemplating bile-likeness) *you can bring something about* (anger reduction) implies *anger is evitable*; however, that implication arguably favors *soft* over *hard* determinism—if it actually does favor either hard or soft determinism, but it is enough here that it is, *prima facie*, logically consistent with either. The inconsistency objection Śāntideva entertains, therefore, does not *obviously* target *hard* determinism; that doctrine is not explicit in Śāntideva's writings. What it explicitly targets is *impersonality* (nonagential bile-likeness) in worldlings versus *personality* (agential bile-unlikeness) in Buddhists.

The Buddhist view of worldlings is that they are particularly overwhelmed by delusion, which implies their ignorance of dependent origination renders them on par with *impersonal* (nonagential) bile, which *cannot* control itself, whereas verse 32 implies those aware of dependent origination *can* control themselves to progress soteriologically. But the idea that knowledge of impersonal causal/volitional forces leads to agential self-regulation and freedom regarding those forces (however impersonal, determined) is arguably and intuitively more of a *soft* than a *hard* determinist idea, although some hard determinists claim that all such abilities are consistent with hard determinism.

It is one thing for data to be consistent with two theories; it is another for one theory to better accommodate that data. All *determinists*—



soft and hard—*agree* that everything traces to impersonal causes predating agents' existence. But they dispute whether our satisfaction of agent-*proximal* conditions—the data for the question of theory-superiority at issue—suffices for responsible agency. Consider this list of agent-proximal conditions the satisfaction of which (by agents) counts here as data:

1. An agent can deliberate about competing choices that represent alternative futures and by selecting one of them she can be the central causal factor that brings about that future.
2. There is a certain harmony between her actions and considered judgments or values.
3. She can alter her values in light of relevant information.
4. She can form effective meta-volitions.
5. She can alter her volitions to have the sort of will she wants.
6. There is a certain agency-nonundermining history to her volitions and meta-volitions.
7. There is a certain mesh between her volitions and meta-volitions.
8. She exhibits self-regulating volitional control.
9. She exhibits counterfactual control over her volitional structure.
10. She would have done otherwise, had she wanted to do otherwise.

11. She would have done likewise, if she wanted to, even if she was able to do otherwise.
12. She exhibits reason-responsiveness.
13. Certain agency-undermining manipulative conditions are absent.

All determinists agree that many often possess such abilities, though some hard determinists might deny agents ever satisfy some such conditions (say, 9-11).<sup>56</sup> But all hard determinists likely accept a “generic hard determinist principle”:

No matter what agent-proximal conditions are satisfied, agents never exhibit free will in the responsibility-entailing sense.

Soft determinists disagree over which conditions constitute responsible agency, but likely accept a “generic soft determinist principle”:

If certain agent-proximal conditions are satisfied, agents exhibit free will in the responsibility-entailing sense.

Thus, Śāntideva’s worldling may be overwhelmed by conditions that defeat her satisfaction of soft determinist responsible agency criteria, but this does not mean that no one ever satisfies them (as Śāntideva’s objection implies, Buddhist aspirants do satisfy them), or that their satisfaction is always insufficient for responsible agency. In replying to the inconsistency objection, Śāntideva arguably employs implicitly soft determinist reasoning about how agents possessing knowledge of causal

---

<sup>56</sup> As I have argued (*Counterfactual*), however, this denial presupposes what I termed “actualism,” the view that only what is actual at some time is possible, and counterfactuals are never actual, so they are not possible. But determinism is an agglomeration of all deterministic laws, but deterministic laws are all counterfactual-supporting generalizations, in which case actualistic determinism is an oxymoron.

operations do not resemble helpless (bile-like) worldlings. This condition may be added to the list of conditions hard and soft determinists agree some of us satisfy but dispute whether such satisfaction constitutes responsible agency:

14. The agent is aware of agent-proximal causal factors and general karmic and/or causal conditions/processes that shape her choice parameters, she can reflect on which choice is dharmic,<sup>57</sup> and she can make and effectively act on that choice, even in the face of phenomenologically powerful dispositional counter-tendencies.

As I have argued elsewhere (“Meditation”), to the extent Buddhist practitioners cultivate meditative discipline, condition 14 becomes increasingly true of them. The same holds for a variety of free-will-related agent-proximal conditions implicit in central tenets of pan-Buddhist doxography:

15. The agent has significantly dharmic views.
16. She has significantly dharmic volitions.
17. She has significant skill in selecting/performing dharmic actions.
18. She can exert appropriately calibrated effort.
19. She has significant one-pointedness skill.
20. She has significant mindfulness skill.

Item 14 is derivable from *Śāntideva*; items 15-20 are from the Eightfold Path. These items are not exhaustive, but are among those a Buddhist

---

<sup>57</sup> The term “dharmic” means “*Dharma*-oriented” (skillful, relative to Buddhism’s prime directive, liberation).

soft determinist would include in her list of conditions satisfaction of which arguably constitutes something like a Buddhist conception of free will. Thus, Śāntideva arguably favors soft over hard determinism.

Refining his analysis of the determinist thread of paleo-compatibilism, Siderits unpacks the idea that tropes are the ultimate-reality-level phenomena that determine our mental states: tropes may be described as abstract particulars, instantiations of—physical or *mental*—universals, “such as particular occurrences of white, sweet, cold . . . desire, attention, etc.” (“Reductionism” 36). Siderits illustrates these particulars with the case of a person eating vanilla ice cream who desires more. One might hope that whereas purely physical items subject the mereological reductionist to the dilemma of infinitely divisible magnitude, because tropes are mentality-involving they are not entirely physical, so they escape this dilemma.<sup>58</sup> But, setting aside the idea that mental states cannot exist apart from their physical instantiations, even putatively purely mental visual images of white ice cream—whether in the form of (at least once removed) mental representations or direct perceptions—have magnitude; so, they are infinitely divisible (they cannot be real atoms), despite being mental.

Siderits claims that because no statement seemingly about tropes can be conventionally true because trope-talk involves ultimate vocabulary, conventional trope-like-talk must be construed as “person-adjectival,” whereas ultimate trope-talk is impersonal and “trope-atomistic” (36). But conventionalese whiteness is not a person-attribute just because it is discussed by speakers of conventionalese who accept

---

<sup>58</sup> This “hope” is not meant to represent the view Siderits describes, but just as a hypothetical reason to think there may be logically possible trope-types that escape the dilemmas facing physical tropes. Understanding Ābhidharma’s *dhammas* in the Sautrāntika way as tropes, there will be mental and physical tropes. For them the latter would involve occurrences of yellow and sweet. For a modern, they might involve things like occurrences of a certain force at a certain point-instant. Siderits (personal communication, February 2012).

personhood;<sup>59</sup> enlightened beings may use whiteness-talk. The whiteness beings perceive is, on my gloss, phenomenal, thus whiteness-talk is conventional. Paleo-compatibilists *could* say that conventionalese about volitions and perceptions is typically attributed to unreal persons, so its person-adjectival elements cannot be true ultimately, and ultimate events to which volitions and perceptions correspond are real but do not correspond to persons, but *valid* trope-talk can be both conventionally and ultimately true.

For Ābhidharmikas, mental states are high-level aggregates of *dhammas*, indefinitely many (trillions) of which constitute a moment of consciousness (Dhamma and Bodhi “Introduction”). Thus, countless individually imperceptible white-trope momentary events (“trope-atoms”) constitute visible white, and countless individually imperceptible desire-trope-atoms constitute felt desire. Any phenomenon that reduces to trope-atoms is conventionally real, but whatever does not reduce is not conventionally real. With “I want more vanilla ice cream,” there are no I-trope-atoms or ice-cream-trope-atoms, but presumably desire-trope-atoms, sweetness-trope-atoms, and white-trope-atoms, and perhaps consciousness-trope-atoms. But saying “ice cream is an illusion but whiteness and sweetness are real” conjures the no-leather-shoes and the regular coffee objections.

---

<sup>59</sup> Siderits thinks I’ve misunderstood his reasoning here because, on his reading, common sense denies that there can be the occurrence of whiteness without some *thing*—a substance—that is white. That, he adds, is also Candrakīrti’s view of conventional truth—and so what Mādhyamaka implies. So even if I (using my gloss) take his words outside the Ābhidharma context that he’s formulated them within, I’ve missed the point. Siderits (personal communication, February 2012). I may be missing the point, but I doubt it is a given that common sense equates whiteness with substance, when students (who are not philosophy majors) discussing Descartes seem to know that red is a color and triangles are three-sided even if we are dreaming or in an illusory (insubstantial) world. Siderits also seems to evidence one of his glosses here, in implicitly equating common sense with conventionalese. Enlightened beings speak conventionalese, but insofar as common sense may be described as the sum of all mankind’s prejudices, enlightened speakers of conventionalese do not possess common sense.

In their “Introduction” to *A Comprehensive Manual of Abhidhamma*,<sup>60</sup> the most extensive canonical account of Buddhist atomism, Dhamma and Bodhi state:

Briefly, the dhamma theory maintains that ultimate reality consists of a multiplicity of elementary constituents called dhammas. The dhammas are not noumena hidden behind phenomena, not “things in themselves” as opposed to “mere appearances,” but the fundamental components of actuality. The dhammas fall into two broad classes: the unconditioned dhamma, which is solely Nibbana, and the conditioned dhammas, which are the momentary mental and material phenomena that constitute the process of experience . . . . It is the dhammas alone that possess ultimate reality: determinate existence “from their own side” (*sarupato*) independent of the mind’s conceptual processing of the data.<sup>61</sup>

The general justification for their view is the Buddha’s claim to have directly perceived the *dhammas* with the penetrating insight of his enlightened mind, a claim confirmed by subsequent enlightened Buddhists and supported by meditative phenomenology, which reveals the ephemeral, pixel-like, micro-level nature of everything in the mind/body field that otherwise appears as a solid, permanent, macro-level object.

Buddhist tropes are experientially homogeneous aggregates, like whiteness, that decompose into parts, down to indefinitely many micro-

---

<sup>60</sup> In the title, “*Abhidhamma*” is Pāli; “*Ābhidharma*” is Sanskrit.

<sup>61</sup> “*Nibbana*” (as they spell it, but technically, “*nibbāna*”) is Pāli for “*nirvāṇa*” (Sanskrit). In this context, it refers to emptiness. (Likewise, what they spell as “*sarupato*” is, technically, “*sarūpato*.”) Unconditioned *dhamma*, emptiness, is ultimately real for *Ābhidharmikas* and on my gloss, but *Ābhidharmikas* include conditioned *dhammas*, momentary phenomena, in ultimate reality; my gloss parses all conditioned phenomena, all form, as conventional, but also accepts that some conventional truths (their conditioned *dhammas*) are also ultimately true.

level *dhammas*/trope-atoms. This appears reasonable initially; recall, however, that whatever can aggregate (even micro-tropes) has magnitude and is divisible. Only what lacks magnitude is indivisible, but nothing lacking magnitude aggregates. Further, anything above the trope-atomic level, such as my *wanting ice cream*, say, “mental state *x*,” *already* involves mereological fiction because there are no *x*-trope-atoms. For although *some* mental states (meditative trances) may be constituted by a homogeneous mass, say, of monadic/identical bliss-atoms, *most* mental states are constituted by a nonmonadic/heterogeneous variety of items only some of which seem composed of smaller quantities of identical atoms.

Mental state *x* is heterogeneous and minimally dyadic: it involves wanting and ice cream wanted (not to mention whiteness and sweetness). Siderits treats *x* as legitimate even though one of its parts, ice cream, does not exist (does not reduce to ice-cream-tropes) and though *x* is not made of *x*-tropes, perhaps because *x*’s parts—wanting, sweetness, and whiteness, say, *a*, *b*, and *c*—reduce one-to-one to *a*-tropes, *b*-tropes, and *c*-tropes, respectively. But why is similar legitimacy denied to person “*p*” even though there are no *p*-tropes that compose *p*, and *p* arguably reduces by a process similar to the one in which *x* does not reduce to *x*-tropes but nonetheless reduces to *a*-tropes, *b*-tropes, and *c*-tropes? That is, *p* arguably equally reduces to a series of mental states, *x*, *y*, *z*, each of which in turn (like *x*) reduce to *a*-tropes, *b*-tropes, and so on. If mental (“*M*”) states are *heterogeneous*, they do not decompose into a mass of *homogeneous* *M*-trope atoms, so *M*-states are mereological fictions. There can be few *ultimate* *M*-states, if any, on the mereological reduction model (such as all-bliss trance states), and fewer ultimate psychological *laws* that govern them, because *M*-laws govern complexly nonmonadic causal relations *between* *M*-states.

Siderits states, “The illusion of incompatibilism may arise . . . by smuggling the concept of a person into the ultimate level . . . or by importing psychological determinism into our conventional talk of persons” (“Beyond” 155-156). But “psychological determinism” is the thesis that *mental-state* events (“that correspond to . . . deliberating and willing”) are determined. A generic principle of *psychological* determinism might be the practical syllogism rule:

If agent A desires z and believes doing y will bring about z,  
then, *ceteris paribus*, A will do y.

Laws of psychological determinism, if any, have this heterogeneous (triadic, quadratic, or increasingly complex) form (involving agents, beliefs, desires, and actions), or something like it, so they too are mereological fictions. Therefore, *contra* paleo-compatibilism, *psychological determinism*—not about micro-level trope-atomic homogeneous/monadic M-states, but macro-level heterogeneous/nonmonadic M-states—cannot be ultimates. Of course, the paleo-compatibilist can again appeal to the conventional/ultimate spectrum for wiggle room, but the more these appeals are made, the more support they provide for my Mahāyāna-based emptiness/form gloss, if not indirect support for Siderits’s own (post-Reductionist) semantic nondualist synthesis (of Reductionism and antirealism).

Thus, for mereological reductionists, psychological determinism involves mereological fiction and is ultimately false, even if true in some nonultimate (conventional) domain. Garfield’s omni-truth-levels model seems tailor-made to the conventional/ultimate (form/emptiness) spectrum, like Dennett’s, for it suggests that there are many level-appropriate vocabulary/phenomena pairings and that all such pairing-levels are valid, but without reductive eliminativism. As he puts it, “let a thousand entities bloom, requiring of each that it genuinely toil and spin, accomplishing some real explanatory work” (Garfield “Nāgārjuna”



512). If so, if they do explanatory work, responsible agents, M-states, their psychophysical constituents, and *dhammas* possess *equal* ontological value.

This line of reasoning is consistent with the sort of semantic nondualist view Siderits espouses as his own in *Persons*, which raises the question why Siderits has not simply offered the semantic nondualist view of free will. I cannot say, but I will speculate. In light of his explanation of why he does take the role of *amicus theoria* for the Buddhist Reductionist view (as an illustration of the philosophical potential of Buddhism), I suspect the answer has to do with the popularity of reductionism in the audience to which Siderits targets this illustration, namely, Western analytic philosophy. Another possibility is that as difficult as it is to grasp the meaning of Buddhist Reductionism, it is even more so with antirealism and semantic nondualism. Thus, perhaps it is wise to simply leave the matter so that any such interlocutors who are drawn in by the illustration may then go on to be drawn into the dialectical progression from Reductionism through antirealism to semantic nondualism. This would effectuate a philosophical bait and switch, but Siderits is explicit in his *amicus theoria* disclaimers. All of this is, admittedly, speculation. However, it bears repeating that it would be nice to hear Siderits's own articulation of the semantic nondualist view of free will.

## Conclusion

Let us conclude our assessment of Siderits's *amicus* briefs on behalf of paleo-compatibilism. By applying Buddhist Reductionism to free will and determinism, paleo-compatibilism accommodates both libertarianism and determinism—two *prima facie* contradictory doctrines. Thus, paleo-compatibilism warrants the classification “semi-compatibilist,” a term previously denoting only the view that determinism is “hard” because it

is incompatible with metaphysical-alternatives-accessing autonomy but “soft” because it is compatible with moral responsibility.<sup>62</sup> But paleo-compatibilism is difficult to comprehensively assess because it stacks a number of puzzling doctrines atop each other, and readers must examine Siderits’s other works to ascertain whether paleo-compatibilism is broadly coherent; arguably, it is not: to the extent his larger work in *Persons* dialectically critiques and absorbs (paleo-compatibilism’s) Reductionism into its larger framework of semantic nondualism, that larger work suggests that paleo-compatibilism is not broadly coherent, though Siderits’s larger framework is. Taken as a stage within that larger framework, it is coherent.

Although paleo-compatibilism raises questions, and some of its details seem at odds with each other, many such discrepancies are functions of external criticisms that shift focus between different nodes along the conventional/ultimate spectrum, rather than indications of internal inconsistencies in paleo-compatibilism. There is no doubt that elements of semantic dualism, mereological reductionism, and trope-theory *do* appear in various forms of Buddhism, especially pre-Mahāyāna Buddhism. Their seeming incompatibility, likewise, may owe more to the complexity of Buddhism than to paleo-compatibilism *per se*. To his credit Siderits seeks to integrate them all under one theoretical umbrella that ought to earn the respect of Western philosophers. Nonetheless, compatibilists and incompatibilists alike might consider paleo-compatibilism *faux-compatibilism*, for the generic reason they equally resist semi-compatibilism despite its offer of an olive branch to both sides, and perhaps also because the opacity of semantic insulation obscures the inconsistency between determinism and indeterminism—its cage compatible character.

---

<sup>62</sup> See Fischer (Way).

Those Western philosophers who do look to Buddhism for support have questions for paleo-compatibilism. How can determinism be ultimate if all nomological relations are minimally dyadic and moments-spanning, which contradicts the Buddhist idea that momentariness undermines the ultimate status of anything moments-spanning? How can there be ultimate “moments” or “instants”—temporal atoms (indivisibles)—if anything with temporal magnitude is infinitely divisible? Ābhidharma posits indefinitely many *dhammas* per blink of an eye, but even if these somehow manage to transcend logic and actually be *magnitude-lacking* indivisibles that can nevertheless aggregate to form *magnitude-possessing* perceptible elements of experience, anything involving more than one extremely small micro-moment—which is everything we experience and know about—does not exist. Any claim about moments-spanning relations or causal processes—sufficient for any perceivable iota—is ultimately *unreal*. The paleo-compatibilist reply is likely simple: this is correct, but do not forget the other part of the paleo-compatibilist explanatory strategy, which is that all these things are conventionally real. And Siderits can always add that the full explanation may be found in his semantic nondualism, where a thousand flowers bloom.

My analysis suggests that determinism and free will are both conventional, so both are robustly compatible (even if located at different nodes along the conventional spectrum). But Sautrāntikas think whatever has causal powers is ultimately real.<sup>63</sup> Because natural kinds are defined by their causal powers, they must be ultimately real. Thus, laws about them, and generalizations about all laws (determinism), must be ultimate as well. This “if causal, then ultimate” formulation supports the causal/counterfactual analysis of an agent-like process or person-series that exhibits the sort of self-regulatory (causal) control over

---

<sup>63</sup> Cozort and Preston (*Buddhist* 55).

itself (autonomy) depicted by satisfaction of some of our (arguably soft-determinist-favoring) items 1-20 (particularly 9-11 and 14). But then both free will and determinism would be ultimately and robustly compatible (in the sense of being true on the same bivalent scale), or both conventionally and equally unproblematic.

All such objections notwithstanding, it is to Siderits's credit that he sticks his neck out to plant fertile ground for humanists to try to salvage free will in the ever-encroaching face of increasingly threatening determinism, by reference to paleo-compatibilism. No complete account of the Buddhist understanding of free will can ignore Siderits's central contribution—the claim that paleo-compatibilism offers a philosophically rich way to understand the issue of free will and determinism—its unresolved implications notwithstanding. As Siderits himself insists, however, this is not necessarily “the” Buddhist view, “the” Ābhidharmika view, or “his” view. Rather, it is just “a” possible view that a Buddhist Reductionist *might* adopt in response to the freedom-determinism problematic. But in spelling out the many ways even one kind of Ābhidharmika model might be developed to this end, Siderits has, undoubtedly, *significantly* raised the level of debate, for any other kind of Ābhidharmika, Buddhist, or non-Buddhist.

Although early-period scholars sought to identify a middle-path between “rigid” hard determinist and “chaotic” indeterminist libertarian extremes, but failed to clearly articulate their positions, Siderits's paleo-compatibilism seeks to salvage elements of both extremes by locating them on different levels of discourse one of which, the conventional (in which persons exist and have free will), reduces to the other, the ultimate (in which there are no persons but only person-series that are entirely determined by impersonal causes). His particular reductionism is mostly limited to pre-Mahāyāna Buddhism and thus it is perhaps unlikely to impress Mahāyānists, compatibilists, or incompatibilists

without further refinements. However, if those refinements develop (perhaps along traditional semi-compatibilist lines and those suggested by the dialectical progression in *Persons*), his strategy seems promising.

In the next article in this series, “Determinism”, I discuss how Charles Goodman embraces hard determinism, arguing that Buddhism rejects *autonomous* agency because it rejects agency or selfhood and because it rejects moral responsibility, which latter presupposes an autonomous self. Siderits and Goodman embrace hard determinism, but in radically different ways, reflecting different reactions to and interpretations of the *anātman* doctrine. In recent-period scholarship (Repetti “Recent”), these divisions run more acutely along doctrinal lines, where scholars relying on Pāli sources mostly accept determinism, but scholars relying on Mahāyāna sources seem to embrace indeterminism. Both groups agree, however, that Buddhism is compatible with free will even in the absence of a real self, a position I call “soft compatibilism” to contrast with “hard incompatibilism,” the view that free will is incompatible with determinism and with indeterminism. In this sense, perhaps Siderits’s semi-compatibilism, in trying to accommodate both sides, is not only consistent with early-period attempts at a middle-path toward free will, but prescient.

## Bibliography

Baker, Alan. “Simplicity,” *The Stanford Encyclopedia of Philosophy*. Online. Available: <http://plato.stanford.edu/entries/simplicity/> (accessed February 21, 2012).

Bird, Alexander and Emma Tobin. “Natural Kinds,” *The Stanford Encyclopedia of Philosophy*. Online. Available: <http://plato.stanford.edu/entries/natural-kinds/> (accessed February 21, 2012).

Caruso, Gregg. *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*. New York: Rowman & Littlefield, 2012.

Cozort, Daniel and Craig Preston. *Buddhist Philosophy: Losang Gönchok's Short Commentary to Jamyang Shayba's Root Text on Tenets*. Ithaca, New York: Snow Lion Publications, 2003.

Dennett, Daniel. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press, 1984.

Dennett, Daniel. *Brainstorms*. Cambridge, MA: MIT Press, 1981.

Dhamma, U Rewata and Bhikkhu Bodhi. "Introduction," *A Comprehensive Manual of Abhidhamma: The Abhidhammattha Sangaha of Acariya Anuruddha*, Bhikkhu Bodhi, editor. Online. Available <http://www.accesstoinsight.org/lib/authors/bodhi/abhiman.html> (accessed February 21, 2012).

Federman, Asaf. "What Kind of Free Will Did the Buddha Teach?" *Philosophy East and West* 60, no. 1 (January 2010), 1-19.

Fischer, John Martin. *My Way: Essays on Moral Responsibility*. New York: Oxford University Press, 2006.

Frankfurt, Harry. "Freedom of the Will and the Concept of the Person," *Journal of Philosophy* 68 (1971), 5-20.

Frankfurt, Harry "Alternate Possibilities and Moral Responsibility", *Journal of Philosophy* 66, no. 23 (1969), 829-839.

Garfield, Jay L. "Nāgārjuna's Theory of Causality: Implications Sacred and Profane," *Philosophy East & West* 51, no. 4 (October 2001), 507-524.

Gier, Nicholas F. and Paul Kjellberg. "Buddhism and the Freedom of the Will: Pali and Mahayanist Responses." In *Freedom and Determinism*, edited

by Joseph Keim Campbell, Michael O'Rourke, and David Shier, 277-304. Cambridge, MA: MIT Press, 2004.

Goodman, Charles. *Consequences of Compassion: An Interpretation & Defense of Buddhist Ethics*. New York: Oxford University Press, 2009.

Gyatso, Geshe Kelsang. *Heart of Wisdom: An Explanation of the Heart Sutra*. New York: Tharpa Publications, 2010.

Harvey, Peter. "'Freedom of the Will' in the Light of Theravāda Buddhist Teachings," *Journal of Buddhist Ethics* 14 (2007), 35-98.

Harvey, Peter. *An Introduction to Buddhist Ethics: Foundations, Values, and Issues*. New York: Cambridge University Press, 2000.

Kasulis, T. P. *Zen Action, Zen Person*. Honolulu, Hawaii: University of Hawaii Press, 1985.

Nāgārjuna. *The Fundamental Wisdom of the Middle Way: Nāgārjuna's Mūlamadhyamakakārikā*, translated by Jay Garfield. New York: Oxford University Press, 1995.

Putnam, Hilary. "Meaning and Reference," *Journal of Philosophy* 70 (1973), 699-711.

Repetti, Riccardo. "Recent Buddhist Theories of Free Will: Compatibilism and Incompatibilism," *Journal of Buddhist Ethics* 18 (2012) (forthcoming).

Repetti, Riccardo. "Buddhist Hard Determinism: No Self, No Free Will, No Responsibility," *Journal of Buddhist Ethics* 18 (2012) (forthcoming).

Repetti, Riccardo. Response to James J. Stewart, "The Question of Vegetarianism and Diet in Pāli Buddhism." Online. Available <http://blogs.dickinson.edu/buddhistethics/2010/05/15/vegetarianism-and-diet-in-pali-buddhism/#comments> (accessed February 24, 2012).

Repetti, Riccardo. "Earlier Buddhist Theories of Free Will: Compatibilism," *Journal of Buddhist Ethics* 17 (2010), 279-310.

Repetti, Riccardo. "Meditation and Mental Freedom: A Buddhist Theory of Free Will," *Journal of Buddhist Ethics* 17 (2010), 24-69.

Repetti, Rick. *The Counterfactual Theory of Free Will: A Genuinely Deterministic Form of Soft Determinism*. Saarbrücken, Germany: LAP Lambert Academic Publishing, 2010.

Schiffer, Steve. "Ceteris Paribus Laws," *Mind* 100, no. 1 (January 1991), 1-17.

Schiffer, Steve. "Physicalism," *Philosophical Perspectives* 4, *Action Theory and Philosophy of Mind* (1990), 153-185.

Searle, John. *The Rediscovery of Mind*. Cambridge, MA: MIT Press. 1992.

Sellars, Wilfred. "Philosophy and the Scientific Image of Man." In *Frontiers of Science and Philosophy*, edited by Robert Colodny, 35-75. Pittsburgh, PA: University of Pittsburgh Press, 1962.

Shantideva and Padmakara Translation Group. *The Way of the Bodhisattva*. Boston, MA: Shambhala, 2008.

Siderits, Mark. Panel. "Contemporary Perspectives on Buddhist Ethics" conference, Columbia University (October 7, 2011). Online Available [http://www.cbs.columbia.edu/buddhist\\_ethics/panel-five.html](http://www.cbs.columbia.edu/buddhist_ethics/panel-five.html) (accessed February 21, 2012).

Siderits, Mark. "Is Reductionism Expressible?" In *Pointing at the Moon: Buddhism, Logic, Analytic Philosophy*, edited by Mario D'Amato, Jay L. Garfield, and Tom J.F. Tillemans, 57-69. NY: Oxford University Press, 2009.

Siderits, Mark. "Paleo-Compatibilism and Buddhist Reductionism," *Sophia* 47, no. 1 (2008), 29-42.



Siderits, Mark. "Buddhism & Free-Will: Buddhist Paleo-Compatibilism," Columbia Society for Comparative Philosophy (April 2, 2007). Online. Available <http://www.cbs.columbia.edu/weblog/2007/04/buddhism-free-w.html> (accessed February 21, 2012).

Siderits, Mark. "Buddhist Paleo-Compatibilism," handout at the Columbia Society for Comparative Philosophy (April 2, 2007). Online. Available <http://www.cbs.columbia.edu/Paleo.pdf> (accessed February 21, 2012).

Siderits, Mark. *Buddhism as Philosophy: An Introduction*. Burlington, VT: Ashgate Publishing Company, 2007.

Siderits, Mark. *Personal Identity and Buddhist Philosophy: Empty Persons*. Burlington, VT: Ashgate Publishing Company, 2003.

Siderits, Mark. "Beyond Compatibilism: A Buddhist Approach to Freedom and Determinism," *American Philosophical Quarterly* 24, no. 2 (April 1987), 149-159.

Thakchoe, Sonam. "The Theory of Two Truths in Tibet," *The Stanford Encyclopedia of Philosophy*. Online. Available: <http://plato.stanford.edu/entries/twotruths-tibet/> (accessed February 21, 2012).

Wallace, B. Alan. "A Buddhist View of Free Will: Beyond Determinism and Indeterminism." In *Neurosciences and Free Will*, edited by Robert Pollack, 59-69. New York: The Earth Institute at Columbia University, 2008.